Canadian Labour Market and Skills Researcher Network

Working Paper No. 86

Universal Early Childhood Interventions: What is the Evidence Base?

Michael Baker
University of Toronto

November 2011

CLSRN is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC) under its Strategic Knowledge Clusters Program. Research activities of CLSRN are carried out with support of Human Resources and Skills Development Canada (HRSDC). All opinions are those of the authors and do not reflect the views of HRSDC or the SSHRC.

Universal Early Childhood Interventions: What is the Evidence Base?

Michael Baker
University of Toronto

Abstract: Universality is a hallmark of Canadian social policy for very young children. The evidence base for these policies is small, non-experimental and offers mixed results. In contrast the evidence base for targeted early childhood interventions is largely experimental and offers strong guidance. Policy makers and advocates often cite the research on targeted programs in support of universal programs, although this is problematic for a number of reasons. Universal programs require a better understanding of the developmental trajectories of more advantaged children. Evidence from the NLSCY suggests there are some potentially important differences in the association of early and later childhood developmental outcomes by family economic resources.

JEL Code: J24, I38, I12
Keywords: Child development, early childhood interventions, maternity leave, preschool

I thank Mike Veall and David Green for giving me the opportunity to deliver the 2011 Innis Lecture. Martin Dooley, David Green and Pierre Fortin provided helpful comments. Alfia Karimova provided outstanding research assistance. I gratefully acknowledge the research support of SSHRC (Grant, #410-2008-0346). Finally, I thank the staff of the Toronto Research Data Centre for their technical support. This paper represents the views of the author and does not necessarily reflect the views of Statistics Canada.

Early childhood is the target of a very active area of social policy in Canada. In recent decades full day kindergarten has been made available to 5 year olds in British Columbia, New Brunswick, Nova Scotia, P.E.I. and Quebec, and to 4 and 5 year olds in Ontario. Starting in the late 1990s Quebec introduced a universal, public child care and early education program available to children from birth. At the end of 2000, Employment Insurance (EI) benefit entitlement for maternity/parental leave was increased from 25 to 50 weeks, and the provinces responded with amendments to their labour standards acts to provide a comparable period of job protection. In 2006 Quebec took over administration of EI maternity/parental leave benefits in its jurisdiction significantly enhancing the level and duration of benefits. Expenditures on the Federal Universal Child Care Benefit are estimated to be near $2.6 billion in 2010/11 (Treasury Board of Canada Secretariat 2010). Looking forward PEI has embarked on a transformation of the care and education of preschool children through the Preschool Excellence Initiative. Ontario's "Pascal report" provides a comparable vision of enriched opportunity for young children in that province (Pascal 2009).

In part these initiatives reflect a trend in developed countries to invest in young children. They are founded on the hypothesis and related empirical investigation that there are strong precursors of poor adult outcomes at young ages. In economics, this argument has been made prominently by James Heckman (e.g., Cunho and Heckman 2010). Some essential elements of the argument are: 1) gaps in cognitive and non-cognitive abilities emerge at early ages, 2) there are critical periods of human development when interventions will have a larger, or perhaps any, effect—some important ones are in early childhood, 3) there are high economic returns to high quality interventions in the lives of disadvantaged children, 4) in contrast, remedial interventions into the lives of disadvantaged adolescents have low returns. Therefore early intervention can be viewed as a tonic for a variety of social ills that plague disadvantaged children as they age. As a result, they can also be viewed as an alternative to transfer programs for adults that are intended to equalize outcomes (Currie 2001).

An important characteristic of Canadian early childhood interventions is that they are universal. Outside of Canada universal programs appear to be neither the norm nor the exception. For example, many of the initiatives in the U.S., such as Head Start are targeted at children defined to be "at risk". In contrast, Scandinavian countries offer models of universal child care and education that dovetail with the conclusion of maternity/parental leave.

The apparent preference for universal programs in Canada is sometimes thought a national trait, especially within the context of health care. However, perhaps one of the most successful Canadian social policies, the Guaranteed Income Supplement (GIS), an income supplement for low income seniors, is targeted, broadly supported and widely credited with changing the trajectory of poverty among the elderly.[1] More to the point, from a developmental perspective I will argue that the case for universal early childhood interventions does not have a strong foundation in evidence.[2] This is perhaps most immediately apparent by the fact the argument for early childhood interventions outlined above speaks to disadvantaged children.

---

[1] See for example Baker et al. (2009).

[2] As a more general point is that there is little evidence of the "optimal" age for a child to move to non-parental care or to start more formal classroom instruction. Universal early childhood interventions come under scrutiny relative to a status quo schooling system, the parameters of which could also be queried.

In the first part of the paper I provide an overview of the evidence base for targeted interventions and a more detailed review of the evidence base for universal programs. The persuasive power of arguments for targeted interventions lies in research that leans heavily on controlled experiments. There is a core of experimental evidence that high quality education and care of "at risk" children can have significant benefit. In contrast, the evidence base for universal interventions relies at best on natural experiments, is much smaller and offers mixed results.

Perhaps as a consequence, policy makers and advocates often use the evidence for targeted programs in the support of universal programs. Drawing on the example of the Perry Preschool Program I provide some arguments why this may not be appropriate.

To move forward there is a clear need for more research.[3] Three key questions are, 1) do interventions shown to improve the lives of disadvantaged children have comparable effects on more advantaged children, 2) what are the impacts of the interventions actually delivered by universal programs, and 3) does the connection between poor childhood and adult outcomes observed for disadvantaged children hold true for those children with better opportunities? The answer to the first two questions lies in the continued study and evaluation of programs around the world that offer universal access. In the last part of the paper I offer some new evidence on the third question tracking the cognitive and behavioural development of the cohort of Canadian children surveyed in the National Longitudinal Survey of Children and Youth (NLSCY).

[3] I note that in an Innis Lecture over a decade ago, Shelley Phipps (Phipps 1999) made a similar call for more research on the well being of Canadian children, although in a different context.

The Evidence Base for Targeted Early Childhood Interventions

Research supporting interventions in the lives of at risk children has been reviewed extensively elsewhere (see Almond and Currie 2010 for a recent review). For the current purpose, I will focus on a core of experimental evidence, which figures prominently in any discussion of these programs.

Perhaps the most studied and cited is the Perry Preschool Program, an experiment conducted in Ypsilanti Michigan in the 1960s involving 123 low income African American children. Fifty eight of these children were randomly assigned to a high quality preschool program at ages 3 and 4, while the remaining 65 received no program. Program participants have been surveyed periodically about their lives, the most recent data from age 40.

The estimated impact of the program on a variety of outcomes is quite remarkable. A summary of some of the findings is presented in figures 1 and 2. For example, relative to the control group, the proportion of program participants that had an IQ of 90 of higher at age 5 was 139 percent higher, that graduated high school was 44 percent higher, that earned at least $20,000 at age 40 was 50 percent higher and than had been arrested 5 or more times by age 40 was 35 percent lower. The treatment group bettered the controls on a variety of outcomes throughout their lives including achievement tests at primary and secondary school ages, employment rates and home ownership at ages 27 and 40.

Estimates of the economic return to the Perry Preschool investment vary widely, some as high as 16 or 17 percent. In a recent reconsideration of the evidence, which accounts for some compromises in the random assignment of children to treatment, Heckman et al. (2010) estimate a statistically significant annual rate of return of 7-10 percent; as they note above the historical return on equities.

Related evidence from other experimental interventions such as the Early Training Project, the Carolina Abecedarian Project and Milwaukee Project generally support these findings (see Blau and Currie 2006 and Almond and Currie 2010 for a review). Carneiro and Heckman (2003) summarize that the evidence from these programs indicates improvements in social skills, motivation, achievement (although not IQ) and reductions in crime and deviant behavior.

A caution often raised about this evidence is the small scale of the programs and the high quality of the education provided. What would result from a larger scale operation funded at more "reasonable" levels? The U.S program Head Start provides an opportunity to address these concerns. This is a large federally funded program that provides preschool education, health care and social assistance to low income families. In 2010 over 900,000 children were enrolled at a cost of over $7 billion dollars, or $7,600 per child (U.S. Department of Health and Human Services 2010a).

While there have been a number of observational studies of the impacts of Head Start (see Currie 2001 for a review) a recent experimental investigation (U.S. Department of Health and Human Services 2010b) is perhaps more definitive. The randomization compares Head Start eligible children provided access to Head Start programs to children who were not, but who enrolled in other programs as their parents decided. Therefore the measurement is of the impact of Head Start relative to other options available to low income children.

The study finds short term improvements for treated children in cognitive skills, health and parenting that largely disappear by the end of first grade. There is also a finding of some longer term benefit to parent-child relationships for those who enter the program earlier (i.e., age 3).

A very recent example of this type of evidence that has garnered a fair amount of public attention is the long run impact of Project STAR, which randomly assigned kindergarten students in Tennessee to classrooms of different sizes. Chetty et al. (2011) examine the impacts of this experiment up to age 27. They find that those children in smaller classes (15 students on average) were more likely to be enrolled in college at age 20. Those with more experienced teachers had higher earnings at age 27.

The lessons of this research are widely appreciated by policy makers and advocates. Their intellectual heft flows directly from the extensive reliance on randomization in this research. They also provide fairly unambiguous guidance for the impacts of high quality interventions in the lives of disadvantaged kids. The only caution is the fade out of the impacts of Head Start. Perhaps the lesson there is the importance of maintaining the high quality of the interventions as scale grows, or that the treatment effect is different for marginally more advantaged children.

The Evidence Base for Universal Early Childhood Interventions

The evidence for universal early childhood interventions is decidedly different. It is smaller in amount and the use of randomization as an identification strategy is rare. Instead policy changes or program introductions are exploited as “natural experiments” to investigate program effects.

From the start it is important to note that the majority of clients of universal programs are not the disadvantaged. Instead they are children whose families inhabit higher levels of the distribution of economic resources. Also, almost by definition, the scale of these programs will be large. Therefore, very high quality programs will necessarily also have a high cost.

I review three major types of programs that alternatively target children who are 1) infants, 2) toddlers and 3) preschoolers. In each case I focus on research that uses policy based variation in program availability rather than studies with a purely observational design.

a) Maternity and Parental Leave

Perhaps the most common "program" for infants in developed countries is maternity/parental leave mandates. These mandates offer employment protection and almost always some level of earnings replacement, although the U.S. is a prominent exception. Of course, the rationales for leave mandates are not exclusively child development. The World Health Organization (2000) recommends that mothers require at least 16 weeks of absence from work after giving birth to protect both their own and their child's health. Nevertheless, the legislation enacting the mandates invariably makes some mention of the importance of the parent/child relationship immediately post birth to child development.[4]

From a developmental perspective the important impacts of a maternity leave mandate are on the amount of maternal and/or paternal care a child receives and on the

---

[4] For example, the American Family and Medical Leave Act, as well as government documents outlining recent reforms in Australia and the U.K. (Commonwealth of Australia 2009, Employment Relations Directorate 2006)

timing of a mother's return to work. While there is a large observational literature on the first effect, there is very little research on the second. However, research from developmental psychology and other fields tells us that a child passes through distinct developmental stages over the first year. Most parents are well aware of the motor and verbal milestones a child achieves in this period. However, there are also thought to be important steps in other dimensions of development such as cognition and socialization. Furthermore, the rapidly expanding field of epigenetics[5] leads us to wonder if some key triggers of genetic expression important to human development are found in the first year. Within this context, the question whether there are better or worse times for a mother to return to work is clearly relevant but is seldom asked in research.

In the absence of maternity leave laws women do not typically return to work right after birth (see Baker and Milligan 2008). In a sample of U.S. mothers who gave birth in 1991, which predates the Family and Medical Leave Act[6], just 11 percent reported working 1 month after birth, 50 percent reported working by 3 months and just under 80 percent report working by 12 months (Brooks-Gunn et al. 2002). These absences are accommodated through a combination of employer provided benefit, unpaid leaves, vacation time and employment termination. So in studying the typical implementation or change in a maternity leave mandate we ask whether there is a developmental advantage to returning to work at 6, 9 or 12 months instead of 3 months.

---

[5] Epigenetics generally refers to processes that cause heritable changes in gene expression without an alteration in DNA sequence. Because environmental factors are thought to be one type of trigger of these processes, there will be an interplay between the expression of an individual's genes and the external context in which they live—between "nature" and "nurture". Epigenetic inheritance can be viewed as a way through which organisms more quickly adapt to changes in their environment. See, for example, Weinhold (2006).

[6] The FMLA provides mothers up to 12 weeks job protected unpaid leave following birth. Individuals working at businesses with 50+ employees are covered by the law.

The majority of evidence from studies of implementations or changes in these laws, summarized in table 1, suggests that there is not. For example, Baker and Milligan 2010 examine a change in Canada's maternity leave mandates that increased the amount of maternal care children received between the ages of 6 and 9 months (on average). They report no impacts on the motor social development of children measured up to the age of 24 months. More generally whether the impacts are short term (Baker and Milligan 2010-Canada), mid term (Baker and Milligan 2011-Canada) or longer term (Dustmann and Schonberg 2008-Germany, Rasmussen 2010-Denmark, Liu and Skans 2010-Sweden) the answer is there appears to be no positive impact on observable children's outcomes. This research spans an array of developmental measures including the motor social development of infants, measures of the cognitive ability and behaviour of preschool children, selective school enrolment, high school completion, GPA, grades and test scores and wages. The leave mandates studied increased the amount of maternal care children received at ages spanning the interval 2 through 15 months. Finally this evidence is from legislative reforms in Canada, Denmark, Germany and Sweden.

While these papers appear to provide unambiguous guidance, they are small in number and there is dissent. Carneiro et al. (2010) report that an increase in maternal care between 8 and 12 months of age in Norway led to a drop in the high school drop out rate of 2.7 percentage points. Looking forward it would be helpful to have evidence on more mid term outcomes and research that provides greater heed to the lessons of other disciplines regarding the different stages of development over the first year.

b) Child Care and Early Childhood Education

As more mothers with small children participate in the labour force, some governments have responded with programs that publicly provide or subsidize child care. While there is a clear "work/life balance" rationale for these initiatives, child development is at least an equally desired goal. As noted above, there is an observational literature that more generally attempts to discover how maternal employment and the attendant non-parental care affects children. The focus here is again those studies that leverage the identification of any effects off of changes to, or the introduction of, public programs.

The example in Canada is Quebec's Family Policy, which introduced publicly regulated, heavily subsidized child care and early education to children aged 0-5 starting in 1997. The evidence from the roll out of this program is decidedly mixed (see table 2). Baker et al. (2008) report negative mean impacts on indices of behavior, while Lefebrve et al. (2008) report negative mean impacts (one fifth of a standard deviation) on Peabody Picture Vocabulary Test (PPVT) scores. Kottelenberg and Lehrer (2011) refine this inference, confirming negative impacts on behavior at the mean, but positive effects for the most disadvantaged children.

Evidence from other countries is also mixed. Gupta and Simonsen (2010) exploit variation across municipalities in Denmark in guaranteed access to centre based preschool (versus family day care). They report that preschool at age 3 has little effect on non-cognitive outcomes at age 7 relative to family day care, but that enrollment in family day care has negative effects for boys born to mothers of low education. More generally longer hours in non-parental care are associated with worse outcomes.

In contrast, exploiting an expansion of Norway's public care system Havnes and Mogstad (2011) report positive impacts. Relative to informal care, the public system delivered higher educational attainment (primarily for children of low educated mothers) and earnings (primarily for girls) at ages 30-40. Further detail of the earnings effect is provided in Havnes and Mogstad (2009). Here they report that the children of low educated parents were the primary beneficiaries of the expansion, while those of high educated parents experienced a loss.

Further evidence from Norway is provided by Black et al. (2011), who exploit discontinuities in the price of childcare to estimate the effect. They report the variation in the price of childcare around the point of discontinuity does not affect utilization but does have a positive impact on academic performance at the junior high level. In this case presumably the mechanism for the positive effect is the extra income the subsidy frees up.

This last study underlines the importance of understanding the counterfactual of research on publicly provided child care. In many instances public care may simply crowd out private care with little impact on mothers' labour supply. For example Fitzpatrick (2010) reports that the roll out of universal pre-Kindergarten in three U.S. states increased preschool enrollment but had little effect on mothers' labour supply.

Overall these few quasi experimental studies do not provide clear guidance. One could construct a case for or against the public option by combining the results from some of the studies surveyed above with the results from selected observational studies.

c) Preschool

The difference between preschool and early childhood education (ECE)/childcare is not completely clear in many settings. By making the distinction I intend to distinguish programs that represent an extension of the primary/secondary school system to younger ages, from those that are more squarely aimed at infants and toddlers. It is the extension of the school system to younger ages that is currently attracting great interest in many provinces.

Most of the evidence on preschool comes from U.S. states (see table 3). Here many of the studies exploit age cutoffs for preschool enrolment in a regression discontinuity design. The very youngest children in a preschool cohort are compared, at the end of their preschool year, to children who by being just a little bit younger had to wait an additional year before enrolling.

One of the more cited examples of this research in Gormley and Gayer's (2005) study of the preschool program in Oklahoma. They report positive impacts on cognitive, motor and language skills for Hispanics and blacks but not for whites. These effects are positively correlated with eligibility for free school lunch. Gormley et al. (2005) provide a similar analysis using a different cognitive measure and report more broad based gains.

Hustedt et al. (2008) use this same research design to investigate the impact of New Mexico's preschool program. They find positive effects on math (50 percent of a standard deviation), vocabulary (25 percent of a standard deviation) and literacy scores (59 percent of a standard deviation). Relative to the population, their analysis sample over represents Hispanics and Native Americans and under represents whites.

Wong et al. (2007) study the impact of relatively high quality pre school programs in five states. In three states the programs are targeted at "at risk" children (MI, NJ, SC), in one it is universal (OK) and in the fifth it is mixed (WV). Effect estimates tend to be positive and are statistically significant in 8 out of 14 cases.

Fitzpatrick (2009) uses a different identification strategy, comparing outcomes across states and cohorts to analyze the impact of the introduction of universal pre-K in Georgia. Her intention-to-treat estimates indicate cognitive gains (6-9 percent of standard deviation for math and 3-7 percent of a standard deviation for reading) for disadvantaged children living in small towns and rural areas.

Providing some historical perspective, Cascio (2009) examines the impact of the introduction of kindergarten in the U.S. exploiting the different timing across states. She reports a small reduction (2.5 percent) in the high school drop out rate and a small increase (1.3 percent) in college attendance for whites but not for blacks.

Overall, the evidence here is of some positive gains most consistently for disadvantaged children. The gains for more advantaged children do not show up as consistently.

What is the Case for Universal Programs?

Given the evidence base for universal programs is so small and the results so often mixed, the case for universal children's programs is often drawn to the research base for interventions into the lives of disadvantaged children. For example, background reports and policy documents in support of the recent full day kindergarten initiatives in British Columbia and P.E.I., and the full day junior (for four year olds) and senior (for five year olds) kindergarten (JK and SK) initiative in Ontario give some prominence to the results of the Perry Preschool Project or similarly targeted programs.[7] Of course, disadvantaged as well as advantaged children are served by universal programs so it makes some sense that part of the evidence base should speak to this group. Nevertheless, there are least two reasons why the use of this evidence in the evidence base for universal programs is potentially problematic.

First, as a basis for predicting the impact of universal programs on disadvantaged children it is important to keep in mind that the experimental early childhood interventions so often cited were expensive and involved high quality treatments. For Perry Preschool, Heckman et al. (2010) reports a program cost per child of $17,759 (2006 USD). At this cost, children aged 3 and 4 received a 2.5 hours per day pre school program delivered 8 months per year for up to two years. It was delivered by teachers with university degrees, certified to teach in elementary, early childhood and special education. There were 4 teachers for every 20-25 students. The parents of these children received weekly home visits of 1.5 hours and participated in monthly group meeting facilitated by program staff (Schweinhart 2004). Masse and Barnett (2002) report an annual average cost per child of the Abecedarian project of $13900 (2002 USD). In this program there were teacher/child ratio's of 1:3 for infants and toddlers and 1:6 for older children (up to age 5). The preschool centres were open 7:30 am-5:30 pm 50 weeks per

[7] In support of the P.E.I. initiative Flanagan (2010) cites the benefits of Perry Preschool as well as the Abecederian Early Childhood Intervention and Head Start. The Ontario Early Learning report cites the results of the "Ypsilanti, Michigan study" (Pascal 2009). Finally, B.C.'s early learning report clearly states that, "evidence in these studies of targeted programs cannot necessarily be applied to broad populations of children" (p. 10), however, this is the only evidence offered of early childhood interventions having specific benefits such as test scores gains and lower rates of teen pregnancy. (Early Childhood Learning Agency 2009).

year. Children received a special individually designed curriculum as well as medical and nutritional services (Campbell and Ramey 1994).

In contrast, universal programs are likely to provide less differentiated instruction in larger groups and at lower costs. For example, the Ontario full day JK and SK initiative is designed for classrooms that average 26 students, one teacher and one early childhood educator. The program runs the standard school day over the school year. Definitive estimates of costs per student are hard to ascertain, but Ontario funds the instructional component of grades JK through grade 3 at $5523.59 per pupil in 2011 (Ontario Ministry of Education 2011). Funding for all purposes per pupil across the entire education system is projected at $10,730 in 2010/11 (Ontario Ministry of Education 2010). Funding per student in the B.C. public education system is $8357 for 2010/11 (British Columbia Ministry of Education 2011). There kindergarten classes are capped at 22 students and have one teacher.

These comparisons raise the important question of whether universal programs deliver the instruction and support to at risk children that have the benefits documented in the experimental research. Ultimately this is an empirical question, but it is clear that prima facie universal programs may provide very different treatments than those in the widely cited targeted interventions.

The other reason why the use of the evidence base for targeted interventions for universal programs raises questions is the possibility of treatment response heterogeneity. One of the lessons of research on the returns to education over past decades is that segments of a population may respond differently to a common treatment (see for example Card 1999 for a summary of some of this research). In application here the question is whether more advantaged children would benefit from the care and instruction provided in the targeted programs as the at risk children did. To a certain extent this question is academic, because as just argued universal programs do not typically deliver the same treatment. But as a matter of evidence it is worth asking whether if universal programs were made more like the targeted ones could we expect similar benefits.

Again the answer to this question is ultimately empirical, but a relevant consideration is differences between the at risk children targeted by Perry Preschool and similar programs and the majority clients of universal programs. The experimental programs were typically targeted at children whose observable characteristics strongly signalled future problems. A sense of how large these problems would be can be gained by focusing on the outcomes of the control group in the Perry Preschool program. Figure 3 provides a view of some of the outcomes of the control group. What is quickly apparent is the children targeted by the program were quite dramatically at risk. Their portfolio of life outcomes contrasts significantly with the average outcomes one would expect to see in a sample of Canadian children. While it does not follow logically that therefore what worked for the children of Perry Preschool would not work for the average Canadian child, one could easily suspect that Perry Preschool children had unique needs that their program specifically addressed. This is not to say that more advantaged children don't have problems, and that their problems wouldn't be attenuated by some intervention. Rather the point is that based on the wide differences in characteristics and outcomes of these two groups, their problems and responses to a common treatment would be different.

Moving Forward

My reading of the evidence base for universal children programs is that there is a lot that we do not know. To move forward we clearly need more evidence on the impacts of programs that serve wide swaths of the population. As documented above the existing evidence is sparse, and in some areas far flung.

Another area in need of more evidence is of how the developmental deficits of more advantaged children play out over the life cycle. Advocates of universal children programs argue that disadvantage is shared by children of all income levels so that targeting will miss the majority of children who need help. For example, in their influential *Early Years Report* Fraser Mustard and Margaret McCain state, "…the majority of children who have problems are not poor children" (McCain and Mustard 1999, p. 98), citing evidence from Orfford et al (1998). Also influential is the work of Willms (2002) who documented that more than 65 percent of "vulnerable children" live in households of the top 3 quartiles of socioeconomic status.[8]

While this evidence does indicate widespread developmental deficits, it is not strictly speaking relevant to the question at hand. The evidence in Mustard and McCain is for children aged 6-16, while Willm's graph covers children up to age 11. Both are cross sectional snapshots. In both cases we do not learn specifically about the problems of children in the earliest years. In both cases we must assume that the problems at older ages had precursors at early ages since the children are not followed longitudinally.

[8] For example, Willms' evidence is cited in Pascal (2009).

The Incidence and Persistence of the Deficits of Early Childhood: Evidence from the NLSCY

In the final part of this talk I provide new evidence on these questions drawing on data from the National Longitudinal Survey of Children and Youth (NLSCY). To my knowledge these are the only nationally representative data that allow the longitudinal study of early childhood in Canada.

While there is a relatively large literature on the longitudinal association of cognitive and behavioural development at younger ages and long term outcomes for other countries, few studies have examined these relationships for Canadian children.[9] Currie and Stabile investigate the relationship between the behavioural scores of children aged 4-11 in wave 1 of the NLSCY on later outcomes. Currie and Stabile (2006) report that higher scores on the hyperactivity measure are negatively related to subsequent scores on cognitive tests and measures of school performance. For example a score at the 90th percentile of the hyperactivity distribution at these ages increases the probability of subsequent grade retention by 6 percent. Currie and Stabile (2009) confirm these results over the first 5 waves of the NLSCY, as well as documenting some negative impacts of the other measures of behaviour.

Dooley and Stewart (2004 and 2007) investigate the relationship between income and cognitive and behavioural development for children aged 4+ using the first 3 waves of the NLSCY and a variety of strategies to address omitted variables bias. Their study

[9] An exception is Romano et al. (2010) who examine the relationship between the cognitive and behavioural markers in the NLSCY at ages 4-5 and grade 3 math and literacy test scores. They find that the cognitive and attention scores were the strongest predictors of subsequent achievement.

of cognitive outcomes (2004) suggests a likely positive impact of income, while their study of behavioral outcomes (2007) does not.

Hoddinott et al. (2002) provide evidence of the persistence of cognitive developmental deficits across the first three waves of the NLSCY for children aged 0-11 in the first wave. They show, for example, that children scoring in the bottom quintile on the Peabody Picture Vocabulary Test (PPVT) at ages 4 and 5 are about 50 percent more likely to be in the bottom quintile of the distribution of math scores at ages 8 and 9 than those who scored in the top quintile of PPVT.

Relative to this research the evidence presented here focuses on cognitive and behavourial problems identified in the pre school years, and how their persistence varies by family income. I also investigate how these problems of early childhood are associated with measures of human capital accumulation such as grade repetition and high school drop out.

NLSCY Data and Empirical Framework

The NLSCY is a panel survey that was conducted biennially between 1994/95 and 2008/09. The original cohort of children aged 0-11 has been followed over the 8 waves of the survey. I focus on children who were aged 0-5 in wave 1 and as old as age 19 in wave 8.

The NLSCY provides a number of measures of the cognitive and behavioural development of young children. At ages 0-5 the cognitive measures are the Motor and Social Development Scale (MSD) based on parent responses for children aged 0-3, and the PPVT administered to children aged 4 and 5. The MSD measures a combination of developmental and pre cognitive skills. The behavioural indices are based on parent responses to questions about their children and measure hyperactivity, anxiety, aggression and prosocial behaviour. A different set of questions are asked for children aged 2-3, and 4-5. In each case except prosocial behaviour, a higher score indicates more problems. A detailed discussion of these measures and possible biases of parent reported outcomes is provided in the online appendix to Baker et al. (2008).

Like all longitudinal data sets the NLSCY has experienced attrition over time. The attrition between waves 1 and 2 is especially large due to reductions in sample size to reduce costs and the abandonment of the sample that was integrated with the National Population Health Survey. In subsequent waves attrition has been more modest. Since the following analysis tracks the children of wave 1 in subsequent surveys it is important to remember that only those remaining in the survey can be followed.

Using these data I offer some answers to three questions. First how do cognitive and behavioural deficits at ages 0-5 play out across the distribution of family income and resources? To investigate this I examine the cross tabulations of the quintiles of the distribution of the developmental indexes and the quintiles of family resources. Two measures of resources are used. First, socioeconomic status (SES) as determined by parents' education and occupational prestige and household income, and provided in the NLSCY.[10] While a variable such as SES is not in common use in economics, it is in other fields and is used here for comparability. Second, a measure of permanent household income defined as the average of household income over the first 3 waves of the survey.

[10] The measure of socioeconomic status provided in the NLSCY is based on the education and occupation of the person most knowledgeable about the child and any spouse and household income.

The second question addressed is how the developmental deficits of early childhood persist over the lifecycle. Cognitive and behavioural scores obtained in later waves are related to the scores obtained at ages 0-5. The behavioural indices available at ages 0-5 are collected until the children are age 11. On the cognitive side, there are a variety of measurements available that differ by age. For children in grades 2-10 there are the results of math tests based on the Canadian Achievement Tests (Canadian Test Centre 1992). At ages 16 and 17 there is the Problem Solving Exercise. Finally at ages 18 and 19 there are literacy and numeracy assessments based on the 2003 International Adult Literacy and Skills Survey.

In each case I begin with a simple account, graphing the progress of children at different quintiles of the wave 1 scores across the quintiles of scores at later waves (ages). Next I present regression adjusted estimates from the equation

(1) $$Score_{i,w} = \alpha + \beta Score_{i,1} + \varphi Income_{i,1} + \delta X_{i,1} + \varepsilon_i$$

where $Score_{i,w}$ is child i's score in wave w, *Income* is the measure of permanent income constructed from waves 1-3, and X are control variables which include province effects, urban residence effects (5 categories), single year dummies for the child's and mothers' ages, mothers' education (4 categories) and foreign birth, dummies for the number of older and younger siblings, a dummy if there is a spouse or partner present and a dummy for whether the child is male. Note that in wave 1 these children are aged 0-5. An additional specification adding the interaction $Score_{i,1}*Income_{i,1}$ is also estimated to see if any relationship between wave 1 and later scores is attenuated or amplified by household economic resources.

As is well known, a causal interpretation of (1) is complicated by omitted variables that affect both wave 1 and wave *w* scores (see, for example, Case et al. 2005 for a discussion). One strategy to address part of the potential bias is to use sibling fixed effects models. To obtain these estimates the following equation is estimated.

$$(2) \qquad Score_{i,f,w} = \alpha + \beta Score_{i,f,1} + \delta X_{i,f,1} + \lambda_f + \varepsilon_i$$

where *f* is the family indicator and λ_f the family fixed effect. Note that the impact of *Income* (as well as many of the elements of *X*) is not identified in this equation as it is common to both siblings.

Sibling differences will control for fixed unobserved variables that vary at the family level. However, this strategy is only available up to wave 5 outcomes as in subsequent waves only one child per family was followed.

The third question investigated is how these signals of developmental deficit at ages 0-5 impact human capital accumulation. The outcomes investigated are grade retention and high school dropout. Models similar to (1) are estimated, now substituting the outcome measures as the dependent variable. Again unobserved heterogeneity across children could bias the estimates and a sibling fixed effect strategy is used to address this possibility for outcomes observable up to wave 5.

All estimates are constructed using the longitudinal weights provided in the NLSCY.

Results

The analysis begins in figures 4 and 5. In figure 4, I graph the proportion of children in the different quintiles of the cognitive score distribution for households in the top and bottom quintile of SES, respectively. The cognitive scores are the MSD index for children aged 3 and younger and the PPVT for children aged 4 and 5. The quintiles are defined by the distribution of scores by single age categories. There is evidence here that children from more disadvantaged families are more likely to have poorer outcomes. Almost 28 percent of these children score in the bottom quintile of the cognitive distribution versus 16 percent of children from top SES quintile families. The story by permanent income is very similar (figure 5). Here 27 percent of children from the lowest income quintile are in the lowest cognitive quintile versus 18 percent of children from the highest income quintile.[11]

An alternative view of this issue is to ask who are the majority of children in the top or bottom behavioural quintile? In the results by SES, over 48 percent of the children in the bottom cognitive quintile are from the bottom two quintiles of SES. Over 42 percent of children in the top cognitive quintile are from the top 2 quintiles of SES.

Figures 6-9 provide complementary information for the hyperactivity, physical aggression, anxiety and prosocial scales for children aged 2-5 by family SES.[12] As noted above, for each behavioural scale except the prosocial scale a higher score indicates a worse outcome. For the purposes of the figures, I have reversed the scores to maintain the same ordering as the cognitive indexes—a higher score for each scale means a better

---

[11] The pattern by current (wave 1) income is similar.

[12] I use SES rather than permanent income in the following graphs to maximize sample sizes in the individual data cells. A large number of children were dropped from the NLSCY after the first wave. This fact together with the regular attrition of the longitudinal sample leads to a significantly lower number of observations when constructing permanent income using data from waves 1-3. SES is based on family characteristics in wave 1. As can be seen in figures 5 and 6 SES and permanent income tell largely the same story.

outcome. As noted below, I return to the original ordering, familiar to NLSCY users, in the regression results.

For hyperactivity—figure 6—there is greater difference by family SES. Children from the top SES families are almost 22 percentage points more likely to be in the top quintile and over 13 percentage points less likely to be in the bottom quintile than children from the lowest SES families. Over 54 percent of the children in the bottom hyperactivity quintile are from the bottom two quintiles of the SES distribution, while over 49 percent of the children in the top quintile are from the top two quintiles of the SES distribution.

The results for anxiety and particularly aggression—figures 7 and 8—display less of a relationship with SES. For example, for aggression roughly 46 percent of the children in the bottom quintile are from the lower two quintiles of SES and 45 percent of the children in the top quintile are from the two upper quintiles of SES.

Finally the results for the prosocial scale, reported in figure 9, contrast with the results for the other measures. Here a higher proportion of children from the top SES quintile are in the bottom quintile of the behaviour score. Just over 38 percent of the children in the bottom quintile of the prosocial score come from the bottom two quintiles of SES.

This analysis provides some support for the argument that targeting might miss a significant number of children with developmental delay. For example, for the cognitive outcomes the incidence of being in the lowest quintile is non-trivial for children in the top SES or income quintile. That said, targeting on the bottom two quintiles of family resources, a slightly wider target than that used in the GIS, would reach almost half of those children in the bottom quintiles of the cognitive distributions and over half of those in the bottom quintile of the hyperactivity distribution.

Simply documenting that higher income children have a non-trivial incidence of cognitive and behavioural problems at preschool ages is a not sufficient case for universal intervention. Also necessary is evidence that a developmental deficit at younger ages means the same thing for advantaged children as it does for disadvantaged children. As noted above much of the enthusiasm for investment in early child development is based in the belief that the developmental deficits of disadvantaged children persist—a fact clearly evident in figure 3 for the Perry Preschool control group. However, generally less is known about the problems more advantaged children face at young ages.

To shed some light on this issue, in figures 10-15 I provide information on the trajectories of children in the top and bottom cognitive and hyperactivity quintiles by the quintile of their family SES status in the first wave.[13] I focus on these two measures because previous research (e.g., Currie and Stabile 2009) and the results below indicate they are the most predictive of the subsequent outcomes of these children. Figure 10 provides the information on the transition from the bottom quintile of the MSD distribution at ages 0-3 to the quintiles of the Math score distribution at ages 12-15. Again a child's position in the distribution is determined relative to the age specific distribution of scores. Here there are very dramatic differences by SES. A child, who is in the bottom quintile of MSD at ages 0-3, has a 0.43 chance of being in the bottom quintile of the math score distribution if s/he is from a bottom quintile SES family. If

[13] I use SES rather than permanent income for these figures to avoid the additional sample attrition that results from constructing the income measure from the first three waves of data. SES is based solely on wave 1 information. Permanent income is used in the regressions reported below.

s/he is from a top SES quintile family the probability is 0.1. It is also clear that the children from top SES quintile families are far more likely to move up the cognitive distribution as they age. This difference in persistence is also reflected across the entire distribution of income. In figure 11 I graph the probability of persisting at the bottom quintile of the cognitive distribution by quintile of SES. A clear SES profile is apparent. This higher SES status of the family the less likely the child remains in the bottom quintile.[14]

A similar story for children who score at the top of the MSD distribution at ages 0-3 is told in figure 12. Children from the top quintile of SES are dramatically more likely to hold their position at the top quintile while children in the bottom quintile of SES are more likely to move down the distribution, many to the bottom quintile.

Corresponding information for the trajectories of hyperactivity is provided in figures 13-15. Here there is more stability and less difference by SES, although the difference in persistence at the bottom quintile remains striking (figure 13). The profile of persistence at the bottom quintile across quintiles of SES (figure 14) reveals that the primary difference is between the top SES quintile and all other quintiles. Finally the transitions of the children initially in the top quintile (figure 15) are quite similar by SES.

This evidence indicates that while the initial incidence of cognitive delay may be non-trivial at all income levels, it is dramatically more persistent for children from the lower SES families than it is for children from the higher SES families. Similarly

---

[14] A caution here is that because MSD is parent reported, the differences by SES reflect something about the parents by SES rather than the children. However a similar SES profile of the probability of persisting in the lowest quintile is observed using PPVT rather than MSD scores as the wave 1 reference. PPVT is a test. Because PPVT is only administered to 4 and 5 year olds, there are not enough observations to support the same graphical analysis as presented for MSD.

children from higher SES families are far more likely to persist at the top quintiles of the cognitive distribution. The story for hyperactivity is different. While the distribution of hyperactivity at ages 2-5 is skewed toward low SES children, it persists quite similarly for all but those from the highest SES quintile.

The story at the mean is provided by the regression results in table 4. Here I report estimates of equation (1) where the dependent variable is the indicated outcome at waves 3, 5 or 7 and the main explanatory variable of interest is the preschool cognitive or behavioural score. I also report the results of the specification where the wave 1 score is interacted with the control for permanent income. Hyperactivity and the other behaviour scores are now coded as in the NLSCY for the regressions, so a higher score means a worse outcome for all but the prosocial scale. As noted above these regressions can only reveal associations as omitted variables may lie behind a relationship between wave 1 and later wave outcomes.

Statistically significant associations of wave 1 MSD and PPVT scores are found with math and Problem Solving scores in later waves. All the scores are standardized (z-scores) so the estimates are interpreted as the "result" of a one standard deviation increase in the relevant wave 1 score. The short term association is about 15 percent of a standard deviation while the longer term association is 10 percent or less. The exception is the association between PPVT and the wave 7 Problem Solving score, which is close to 20 percent of a standard deviation. All in all these mean associations are relatively modest given the thought experiment of raising the wave 1 outcome by a full standard deviation.

Note that in the specification with the interaction of the wave 1 score and permanent income, in three of six cases a larger association of wave 1 and later outcomes is indicated for lower income children although the estimate of the interaction effect is not statistically significant.

The results for hyperactivity consistently show a statistically significant association between wave 1 and later wave outcomes. The magnitude of these results is harder to interpret—strictly speaking the “result” of a one point increase in the wave 1 score. At wave 3 the association gets smaller with permanent income.

Associations of similar magnitude are observed for the anxiety score. Here however the interaction with permanent income is either not statistically significant or of an unexpected sign (i.e., wave 5).

Finally the results for the aggression and prosocial scores indicate associations that are about half the magnitude of those for hyperactivity and anxiety. Again the estimates of the interactions with permanent income are not very consistent across waves.

Overall these regressions show statistically significant, although in some cases modest, associations at the mean between developmental indicators measured at preschool and older ages. In general the interactions with income are not statistically significant. As indicated in the figures much of the action for these scales is away from the mean.

In the final part of the analysis I investigate how the wave 1 scores are related to measures of human capital accumulation. I investigate the relationship between the behavioural and cognitive scores and grade retention and high school dropout. I also examine the association between wave 1 behavioural scores and subsequent academic performance. Where possible I use sibling fixed effect models in an attempt to estimate causal relationships between the outcomes.

The analysis begins in table 5 where I report the estimated associations between the behavioural scores at ages 2-5 and the measures of grade retention and high school dropout by the indicated wave of the data. Notable here is that almost all the estimated coefficients are very small and statistically insignificant. There does not appear to be any OLS relationship between these measures and subsequent progress in school.

This conclusion contrasts with Currie and Stabile (2009) who report statistically significant associations between hyperactivity scores and grade retention. However, a potentially important difference between those results and the estimates in table 5 is that Currie and Stabile's wave 1 scores are for children aged 4-11. It is possible that the pre school scores used here simply do not mark school performance as scores at older ages do.

In table 6 are the associations between the wave 1 behaviour scores and subsequent math scores and Problem Solving scores. For hyperactivity the mean relationship is consistently negative,[15] small and almost always statistically insignificant. Note that the math outcome is a z-score so the parameter is the standard deviation change associated with a one point increase in the wave 1 hyperactivity score. Once the interaction with permanent income is specified there is a more consistent statistically significant association, larger at low income levels and attenuated with income.

The results for the other behaviour scores are typically statistically insignificant and the parameter estimates switch signs across waves in many instances. There is not

[15] Recall that the behavioural scores are now coded as in the NLSCY, so a higher hyperactivity score, which is a worse outcome, is associated with a lower math or Problem Solving score.

very compelling evidence here that these scores have a substantial association with these measures of human capital accumulation.

In table 7 I report some sibling fixed effects estimates of the relationship between hyperactivity and subsequent math scores. I focus on hyperactivity because it is this behaviour that has the most consistent association with math scores in table 6. It should be noted that the fixed effects estimates are based on small numbers of sibling pairs—77 for the wave 3 results and 350 for the wave 5 results. This is because of the age restriction to children 2-5 in wave 1. As a result this analysis should be viewed as largely exploratory.

The first column of results is the full sample OLS estimates as in table 6, and in the next column are the OLS estimates for the sibling sample. Focusing on the wave 5 results, with the higher number of sibling pairs, the two estimates are quite similar although the standard error is much larger in the sibling sample. The third column of results is the fixed effect estimate. It is more than twice the magnitude of the OLS estimates and statistically significant. A larger fixed effect estimate is in line with the results reported in Currie and Stabile (2009), although the estimate here is about half the size of their fixed effect estimate using hyperactivity scores at later ages. This suggests a causal relationship between preschool hyperactivity and subsequent cognitive performance, in line with the results of studies for older children.

The last three columns of results are for the specification with the interaction between the wave 1 score and permanent income. Here none of the estimates from the sibling sample are statistically significant and the point estimates of the interaction effect are very small.

In table 8 are estimates of the relationship between MSD scores at age 0-3 and subsequent grade retention and high school dropout. Both OLS and sibling fixed effects estimates (where possible) are reported. In the full sample there is fairly robust evidence that a one standard deviation increase in MSD is associated with a 1-2 percentage point reduction in the probability of repeating a grade, and the relationship is somewhat stronger for low income children. Again focusing on the wave 5 results (594 sibling pairs), the sibling sample estimates suggest a causal mean relationship between MSD and grade retention. They compare well to the full sample result, although the standard error gets much larger in the fixed effect estimation. Once the interaction between the wave 1 score and income is added the sibling sample results show greater variability.

In table 9 are corresponding estimates for PPVT. Because PPVT is administered only to children ages 4 and 5 in wave 1, there are not a sufficient number of siblings to construct the fixed effects estimates. To address this problem, siblings' scores from wave 2 and 3 are used. Therefore the key explanatory variable is the sibling difference in PPVT scores in waves 1-3.[16] Note that the measure of permanent income is constructed from waves 1-3 and therefore will be notionally relevant to both siblings' scores. However, these fixed effects estimates do not hold household characteristics constant in the same way as in table 8.

The full sample OLS results again show a consistently statistically significant association between PPVT and both grade retention and high school dropout. At the mean, a one standard deviation increase in PPVT reduces grade retention by up to 3 percentage points and high school dropout by just over 2 percentage points. Also similar

[16] Waves 1 and 2 for the wave 3 results.

to the MSD results, a larger association is indicated for low income children. The wave 5 fixed effects estimates (based on 502 sibling pairs) indicate the mean relationship with grade retention may be causal and quite similar in magnitude to the corresponding fixed effect estimate for this wave for MSD (table 8). However, once again adding the interaction with permanent income leads to unstable and statistically insignificant results.

Conclusions

Universality is a hallmark of Canadian social policy for very young children. The case for these interventions is often made drawing on evaluations of programs for at risk children. Perry Preschool is a high profile example of these programs, which offer random assignment and fairly unambiguous guidance that early, high quality education and care of disadvantaged children can have substantive and long lasting results. In contrast direct evidence of the impact of universal programs is relatively scare, is non-experimental and offers mixed results.

I argue that using the evidence from targeted programs for the promotion of universal ones is problematic for a number of reasons. First, the widely cited targeted programs offered levels of education and care that are typically not replicated in programs for all children. Therefore it is not clear that universal programs can deliver similar benefits to children at risk. Second, even if universal programs could offer similar levels of care and education it is not clear that more advantaged children would benefit to the same degree as the less advantaged. The children targeted in programs such as Perry Preschool were quite dramatically at risk.

To better understand the potential impacts of universal programs we need more evaluations of universal programs already in place, and more research on the developmental trajectories of advantaged children. In the last section of this talk I provide evidence on the developmental trajectories of Canadian children from the NLSCY.

Consistent with the arguments of many policy makers and advocates, NLSCY data indicate that the incidence of preschool cognitive and behavioural delay, while higher for low income children, is non-trivial even at the highest quintile of family income. However, the persistence of preschool markers of cognition at older ages differs dramatically by family SES. Low SES children are far more likely to persist in the lower quintiles of cognitive outcomes while high SES children are far more likely to persist at the higher quintiles. The persistence of behavioural scores is more equal, but these scores in turn appear to have less association with measures human capital accumulation in later childhood. In contrast early cognitive delay has a statistically significant association with grade retention and high school dropout, that sibling fixed effect estimation suggests is causal.

These results only scratch the surface of this topic. Nevertheless, the findings do indicate some important differences in the association between early and later childhood developmental outcomes by family income and SES. Unfortunately the NLSCY has come to an end, and so it will not be possible to investigate how the early development outcomes of Canadian children map into post secondary schooling and labour market entry. That said, a compelling case for *universal* early childhood interventions requires more research on the consequences of, and possible remedies for, the early developmental outcomes of more advantaged children.

References

Almond, Douglas and Janet Currie (2010) “Human Capital Development Before Age Five” NBER Working Paper No. 15827.

Baker, Michael and Kevin Milligan (2008), “How does job-protected maternity leave affect mothers’ employment?” *Journal of Labor Economics*, Vol. 26, No. 4, pp. 655-691.

Baker, Michael and Kevin Milligan (2010), “Evidence from maternity leave expansions of the impact of maternal care on early child development,” *Journal of Human Resources*, 45, Winter, 1-32.

Baker, Michael and Kevin Milligan (2011), “Maternity Leave and Children’s Cognitive and Behavioral Development,” NBER Working Paper 17105.

Baker, Michael, Jonathan Gruber, and Kevin Milligan (2008), “Universal Childcare, Maternal Labor Supply, and Family Well-being,” *Journal of Political Economy*, Vol. 116, No. 4, pp. 709-745.

Baker, Michael, Jonathan Gruber, and Kevin Milligan (2009) “Retirement Income Security and Well-Being in Canada,” NBER Working Paper 14667

BC Ministry of Education, (2010), “News release 2011EDUC0015-000213”.

Berrueta-Clement, John R., Schweinhart, Lawrence J. Barnett. W Steven, Epstein, Ann S. and David P. Weikart, (1984) *Changed Lives: The Effects of the Perry Preschool Program on Youths through age 19*, Ypsilanti, MI: High/Scope Press.

Black, Sandra E., Devereux, Paul J., Løken, Katrine V. and Kjell G. Salvanes (2011), “Does Money Matter? The Effect of Child Care Subsidies on Academic Performance”, Norwegian School of Economics and Business Administration

Blau, David and Janet Currie (2006). “Pre-School, Day Care, and After-School Care: Who's Minding the Kids?,” *Handbook of the Economics of Education*, Elsevier.

Campbell, F. A. & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, 65, 684-698.

Canadian Test Centre, Educational Assessment Services (1992) *Canadian Achievement Tests*, 3rd edn. Canadian Test Centre, Educational Assessment Services, Ontario, Canada

Card, David, “The Causal Effect of Education on Earnings” in Ashenfelter, O.C. and D. Card, editors, *Handbook of Labor Economics*, New York: Elsevier, 1999, volume 3A, 1801-1863.

Carneiro, Pedro and J. Heckman, "Human Capital Policy," in Inequality in America: What Role for Human Capital Policies? ed. J. Heckman and A. Krueger (Cambridge, MA: MIT Press, 2003).

Carneiro, Pedro., Loken, Katrine V. and Kjell G. Salvanes, (2010) "A Flying Start? Long Term Consequences of Maternal Time Investments in Children During Their First Year of Life", IZA Discussion Paper No. 5362

Cascio, Elizabeth U., (2009), "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing Kindergartens into Public Schools", NBER Working Paper No. 14951.

Case, Anne, Fertig, Angela and Christina Paxson (2005) "The Lasting Impact of Childhood Health and Circumstance", *Journal of Health Economics* 24 365-389.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach and Danny Yagan, (2011), "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence From Project STAR", *Quarterly Journal of Economics*, 126(4).

Commonwealth of Australia (2009), *Australia's Paid Parental Leave Scheme*, CanPrint Communications Pty Ltd.

Cunha, Flavio and James J. Heckman, (2010) "Investing in Our Young People", NBER Working Paper No. 16201, July.

Currie, Janet, (2001), "Early Childhood Intervention Programs: What Do We Know?," *Journal of Economic Perspectives*,15(2), 213-238.

Currie, Janet and Mark Stabile, (2006), "Child Mental Health and Human Capital Accumulation: The Case of ADHD", *The Journal of Health Economics*, 25(6), 1094-1118.

Currie, Janet, and Mark Stabile, (2009) "Mental Health in Childhood and Human Capital", in Jonathan Gruber (ed.) *An Economic Perspective on the Problems of Disadvantaged Youth* Chicago: University of Chicago Press for NBER).

Dooley, Martin and Jennifer Stewart, (2004), "Family Income and Child Outcomes in Canada", *The Canadian Journal of Economics,* 37(4), 898-917

Dooley, Martin and Jennifer Stewart (2007), "Family Income, Parenting Styles and Child Behavioural-Emotional Outcomes" *Health Economics* 16(2) 2007:145-162.

Dustmann, Christian, and Uta Schonberg (2008), "The Effect of Expansions in Leave Coverage on Children's Long-Term Outcomes," Bonn: Institute for the Study of Labor

Working Paper No. 3605.

Early Childhood Learning Agency (2009), “Expanding Early Learning In British Columbia For Children Age Three to Five” accessed at http://www.bced.gov.bc.ca/early_learning/pdfs/ecla_report.pdf on May 17, 2011.

Employment Relations Directorate (2006), *The Maternity and Parental Leave (Amendment) Regulations 2006 and the Paternity and Adoption Leave (Amendment) Regulations 2006, Full Regulatory Impact Assessment*, London: Department of Trade and Industry, June.

Fitzpatrick, Maria D. (2008) “Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement,” *The B.E. Journal of Economic Analysis and Policy* 8(1) (Advances), Article 46.

Fitzpatrick, Maria D., (2010) “Preschoolers Enrolled and Mothers at Work? The Effects of Universal Pre-Kindergarten." *Journal of Labor Economics* 28(1): 51-85.

Flanagan, Kathleen, (2010), “The Early years Report: Early Learning in PEI: An Investment in the Island’s Future”, PEI Department of Education and Early Childhood Development

Gormley, William T. and Ted Gayer (2005) “Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program”, *Journal Human Resources* 40: 533-558.

Gormley Jr., William T., Ted Gayer, Deborah Phillips, and Brittany Dawson, (2005) " The Effects of Universal Pre-K on Cognitive Development", *Developmental Psychology*, 41(6) 872–884

Gupta, Naanita Datta and Marianne Sinomsen, (2010) “Non-cognitive child outcomes and universal high quality care”, *Journal of Public Economics*, 94, 30-43.

Havnes, Tarjei and Magne Mogstad, (2009) “Is Universal Child Care Leveling the Playing Field? Evidence from Non-Linear Difference-in-Differences” IZA Discussion Paper No. 4978

Havnes, Tarjei and Magne Mogstad, (2011) “No Child Left Behind: Subsidized Child Care and Children’s Long-Run Outcomes”, *American Economic Journal: Economic Policy 3 : 97–129*

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev and Adam Yavitz, (2010) “The rate of return to the HighScope Perry Preschool Program”, *Journal of Public Economics* 94 114–128.

Hoddinott, John, Lethbridge, Lynn. and Shelley Phipps. (2002). Is History Destiny? Resources, Transitions and Child Education Attainments in Canada. Ottawa: Human Resources Development Canada.

Hustedt, Jason T., W. Steven Barnett, Kwanghee Jung, and Alexandra Figueras (2008), "Impacts of New Mexico PreK on Children's School Readiness at Kindergarten Entry: Results from the Second Year of a Growing Initiative", The National Institute for Early Education Research Rutgers University

Kottelenberg, Michael and Steven F. Lehrer, (2011) "Reinvestigating Who Benefits and Who Loses from Universal Childcare in Canada", Queen's University.

Lefebvre, P., P. Merrigan and M. Verstraete (2008). "Child Care Policy and Cognitive Outcomes of Children: Results from a Large Scale Quasi-experiment on Universal Childcare in Canada," CIRPÉE - Working Paper: 08-23; DR/WP UQAM http://www.cirpee.org.

Liu, Qian and Oskar Nordstrom Skans (2010) "The Duration of Paid Parental Leave and Children's Scholastic Performance," *The B.E. Journal of Economic Analysis & Policy*, 10(1)_ (Contributions), Article 3.

Masse, Leonard N. and W. Steven Barnett, (2002) "A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention", National Institute for Early Education Research (NIEER), Rutgers, The State University of New Jersey.

McCain, Margaret and Fraser Mustard, (1999), *The Early Years Report*, Toronto: Ontario Children's Secretariat.

Ontario Ministry of Education, (2010), "Fact Sheet: Grants for Student Needs: 2010-11 Highlights", " accessed at http://www.edu.gov.on.ca/eng/document/nr/10.03/bg0326.html on May 19, 2011

Ontario Ministry of Education, (2011), "Technical Paper 2011-12" accessed at http://www.edu.gov.on.ca/eng/funding/index.html on May 19, 2011.

Offord, D., Kraemer, H., Kazdin, A., Jensen, P., & Harrington, R. (1998) "Lowering the burden of suffering from child psychiatric disorder: Trade-offs among clinical, targeted and universal interventions". *Journal of American Academic Child and Adolescent Psychiatry.* 37:7. July 1998: 686 - 694

Pascal, Charles, (2009) "With Our Best Future in Mind", A Report to the Premier from the Special Advisor on Early Learning, accessed at http://www.ontario.ca/ontprodconsume/groups/content/@onca/@initiatives/documents/document/ont06_018899.pdf on May 17, 2011.

Phipps, Shelley, (1999), "Economics and the Well-Being of Canadian Children" *Canadian Journal of Economics,* 32(5), 1135-1163

Treasury Board of Canada Secretariat, (2010), "2011-12 Estimates, Parts I and II, The

Government Expenditure Plan and The Main Estimates" accessed at http://www.tbs-sct.gc.ca/est-pre/20112012/p2-eng.asp on May 10, 2011.

Rasmussen, A.W., (2010) "Increasing the length of parents' birth-related leave: The effect on children's long-term educational outcomes", *Labour Economics*, 17 (1), 91-100.

Romano, Elisa, Babchishin, Lyzon, Pagani, Linda S. and Dafna Kohen, (2010), "School Readiness and Later Achievement: Replication and Extension Using a Nationwide Canadian Survey", *Developmental Psychology*, 46(5) 995–1007.

Schweinhart, Lawrence J., (2004) "The High/Scope Perry Preschool Study Through Age 40, Summary, Conclusions and Frequently Asked Questions", Ypsilanti, MI: High/Scope Press.

Treasury Board of Canada Secretariat, (2010), "2011-12 Estimates, Parts I and II, The Government Expenditure Plan and The Main Estimates" accessed at http://www.tbs-sct.gc.ca/est-pre/20112012/p2-eng.asp on May 10, 2011.

Weinhold B, (2006) "Epigenetics: The Science of Change". *Environmental Health Perspectives* 114(3): doi:10.1289/ehp.114-a160

World Health Organization. (2000). Health Aspects of Maternity Leave and Maternity Protection. Statement to the International Labour Conference, June 2, 2000. World Health Organization: Geneva.

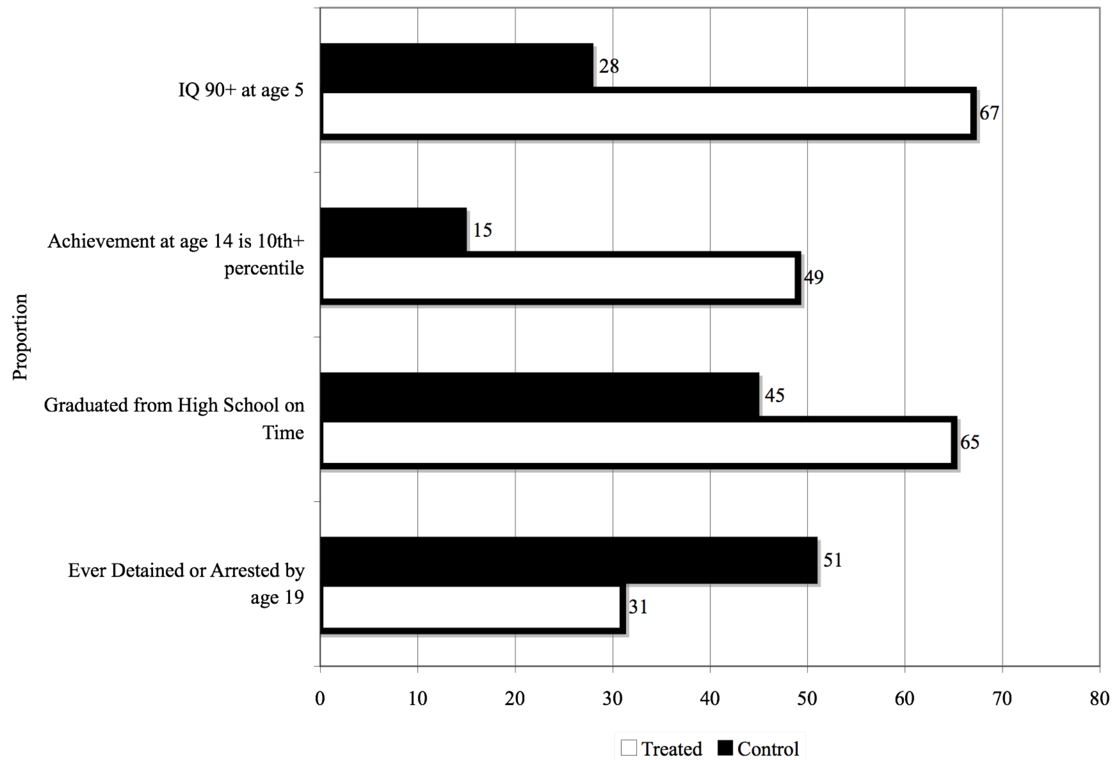
U.S. Department of Health and Human Services (2010a) "Head Start Program Fact Sheet", accessed at http://www.acf.hhs.gov/programs/ohs/about/fy2010.html on May 11, 2011.

U.S. Department of Health and Human Services, (2010b). *Head Start Impact Study. Final Report*. Washington, DC.

Willms J. Douglas (ed.), (2002), *Vulnerable Children*, Edmonton AB: University of Alberta Press.

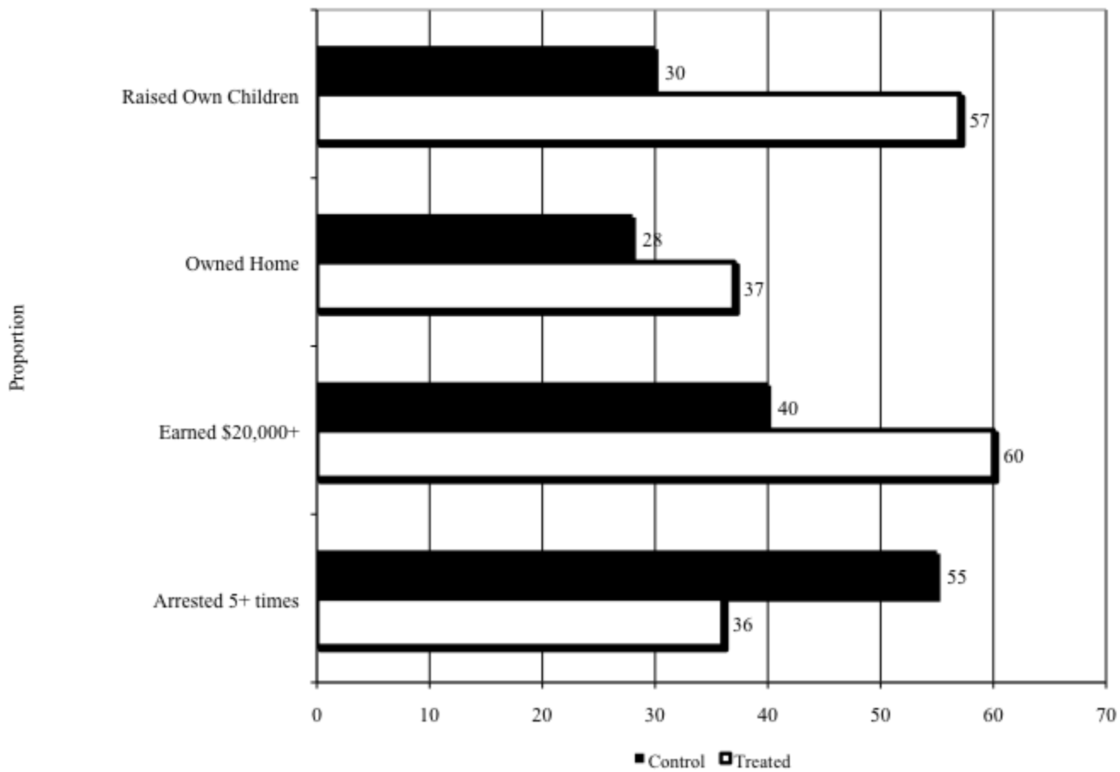
Wong, Vivian C. Thomas D. Cook, W. Steven Barnett and Kwanghee Jung (2007), "An Effectiveness-based Evaluation of Five State Pre-Kindergarten Programs using Regression-Discontinuity", New Brunswick, NJ: National Institute for Early Education Research

Figure 1: Outcomes of the Perry Preschool Program at younger ages



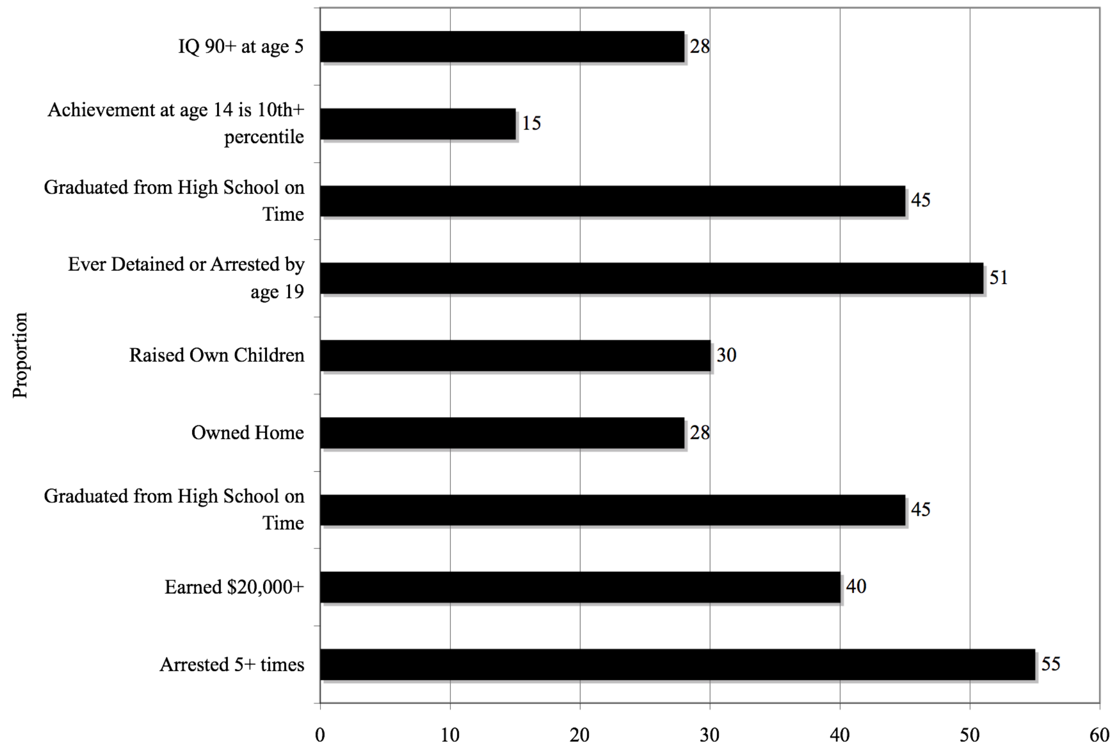
Source: Berrueta-Clement (1984) and Schweinhart (2004)

Figure 2: Outcomes of the Perry Preschool Program at older ages



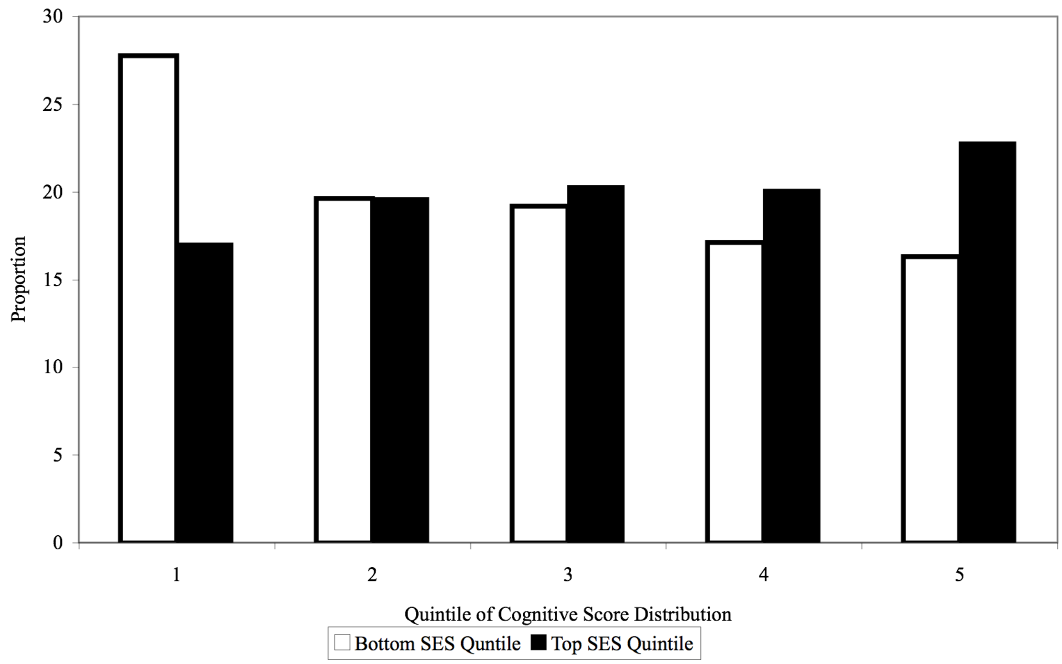
Source: Schweinhart (2004)

Figure 3: Outcomes of the Perry Preschool Program control group



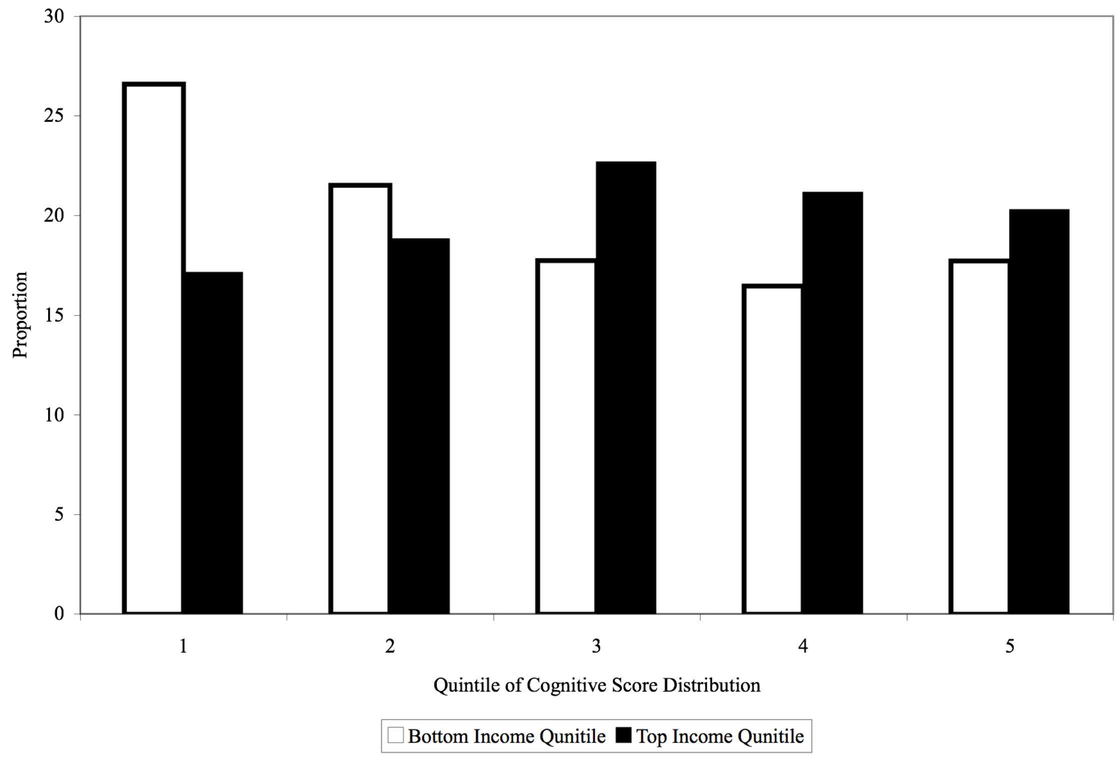
Source: Berrueta-Clement (1984) and Schweinhart (2004)

Figure 4: The Distribution of cognitive scores for children aged 0-5 from the top and bottom quintiles of family SES



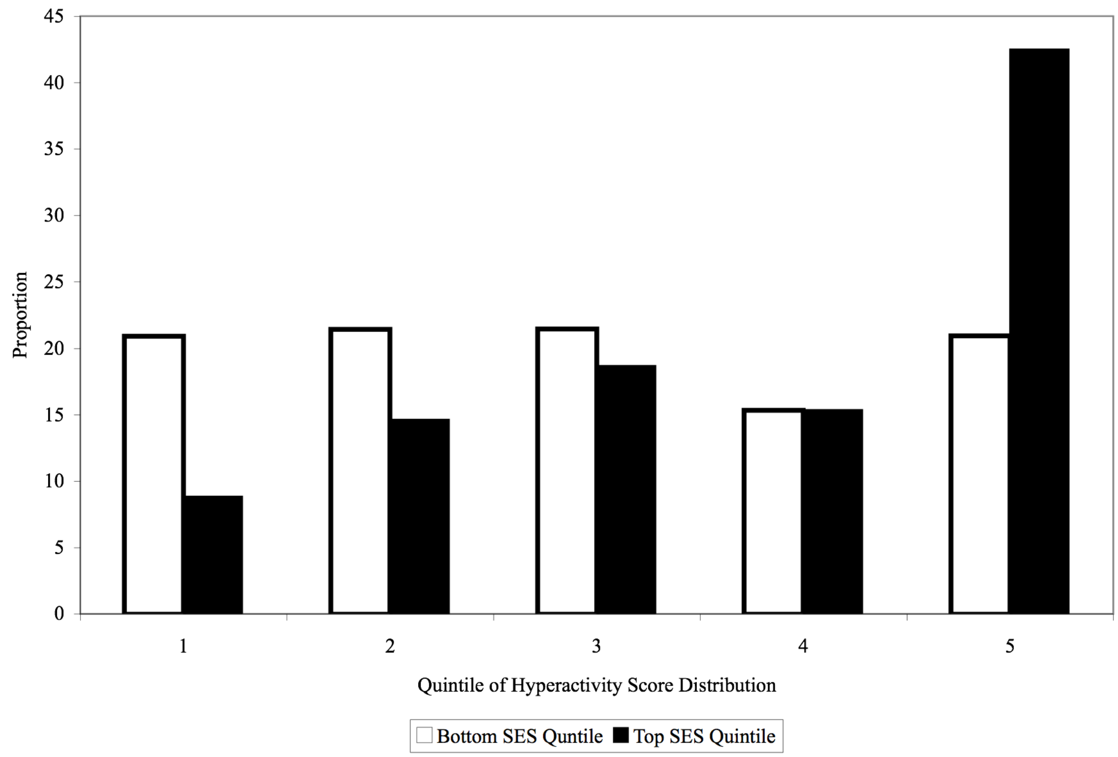
Source: Author's calculations from the NLSCY.

Figure 5: The distribution of cognitive scores for children aged 0-5 from the top and bottom quintiles of family "permanent" income



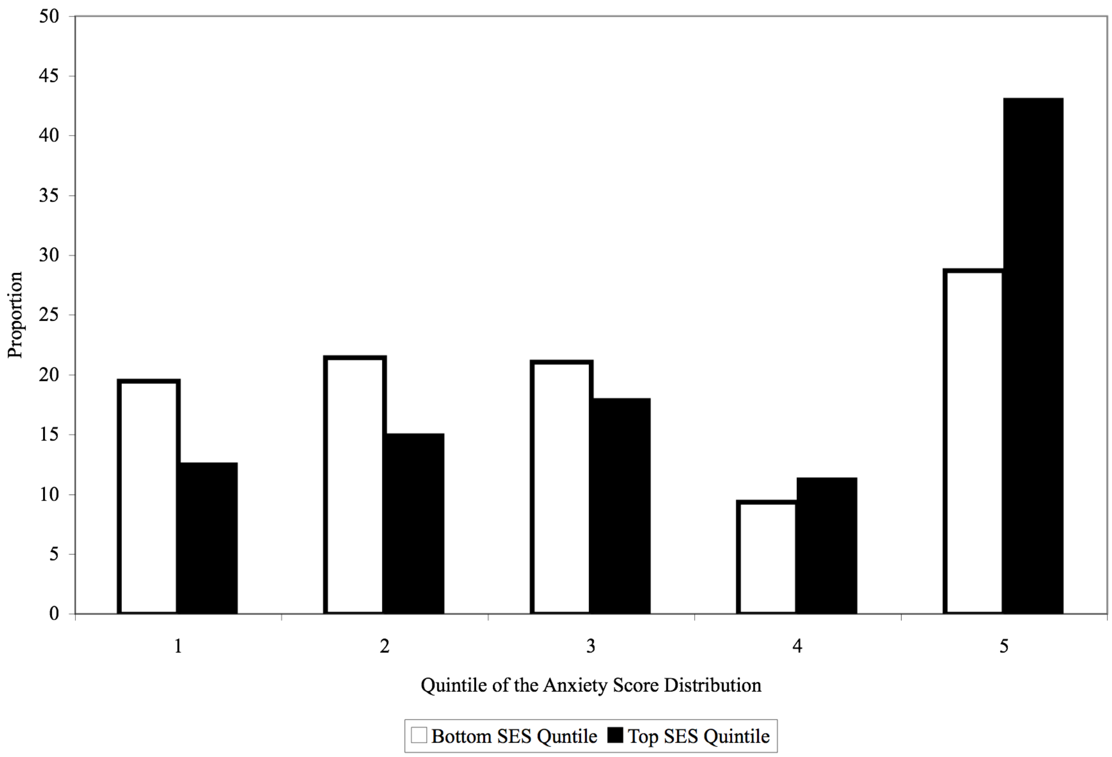
Source: Author's calculations from the NLSCY. Permanent income is constructed by averaging family income over waves 1-3.

Figure 6: The distribution of the hyperactivity scale for children aged 2-5 from the top and bottom quintiles of family SES



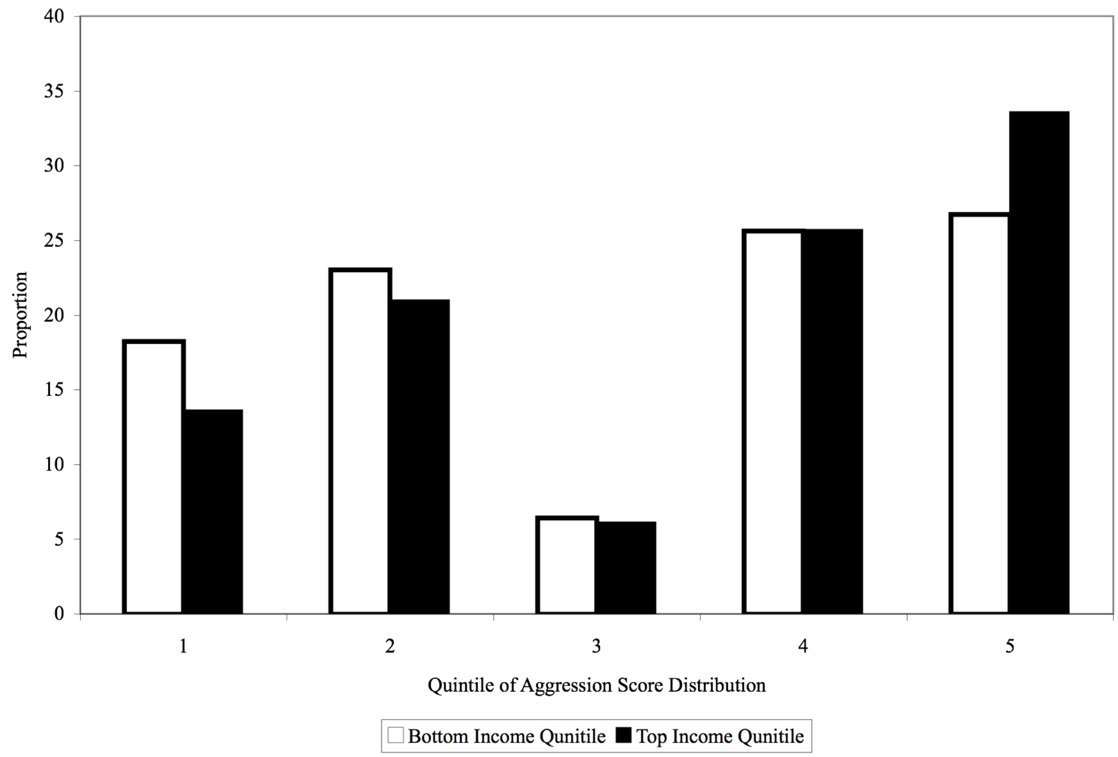
Source: Author's calculations from the NLSCY. Note the ordering of the hyperactivity score is the reverse of the original data—a higher score means a better outcome.

Figure 7: The distribution of the anxiety scale for children aged 2-5 from the top and bottom quintiles of family SES



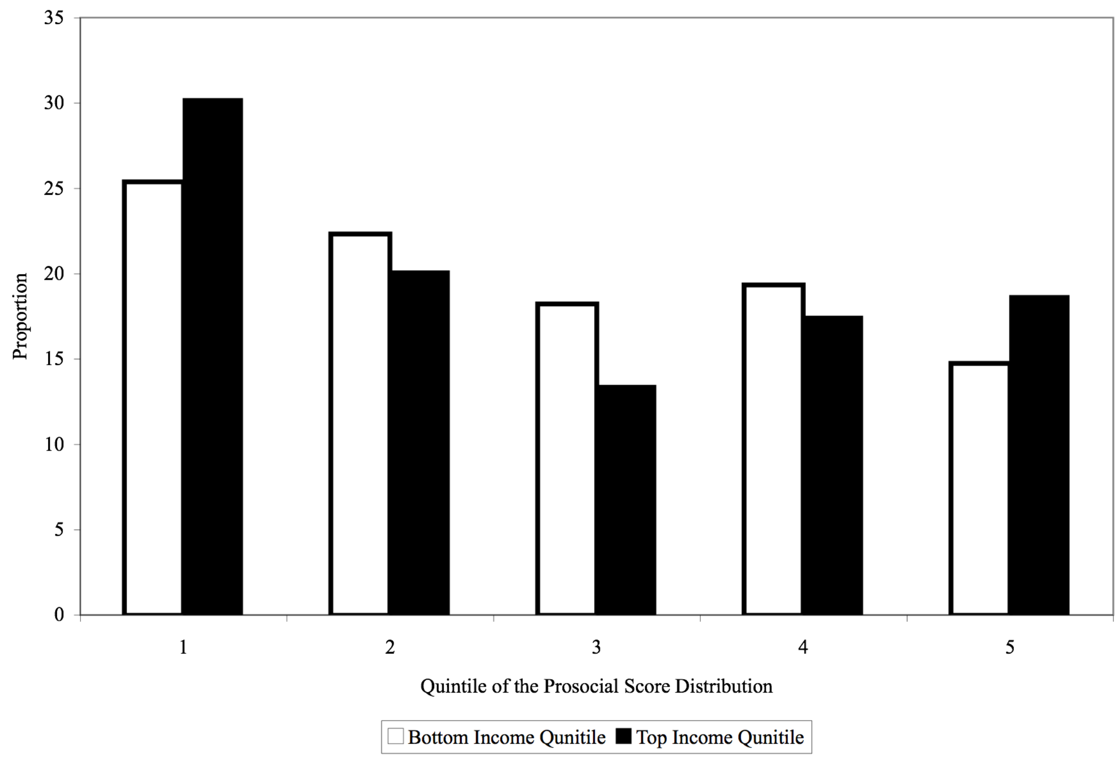
Source: Author's calculations from the NLSCY. Note the ordering of the anxiety score is the reverse of the original data—a higher score means a better outcome.

Figure 8: The distribution of the aggression scale for children aged 2-5 from the top and bottom quintiles of family SES



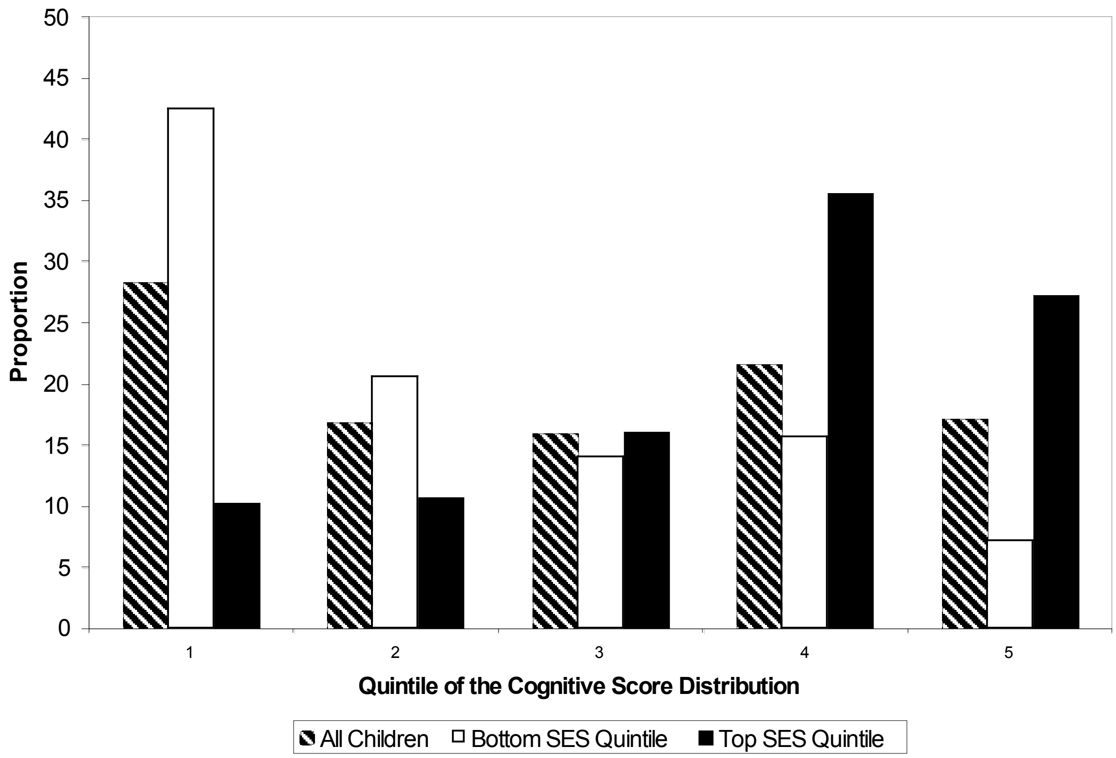
Source: Author's calculations from the NLSCY. Note the ordering of the aggression score is the reverse of the original data—a higher score means a better outcome.

Figure 9: The distribution of the prosocial scale for children aged 2-5 from the top and bottom quintiles of family SES



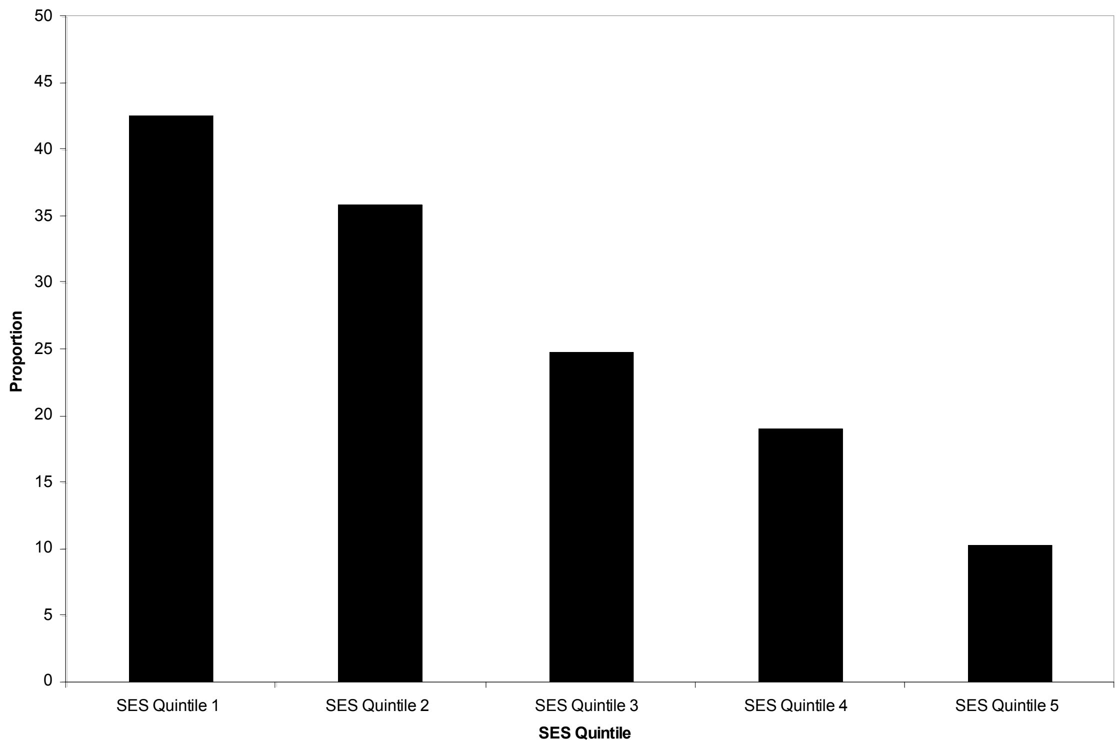
Source: Author's calculations from the NLSCY.

Figure 10: Transition probabilities from the bottom quintile of the cognitive score distribution at ages 0-3 to the quintiles of the cognitive score distribution at ages 12-15, by quintile of family SES at ages 0-3



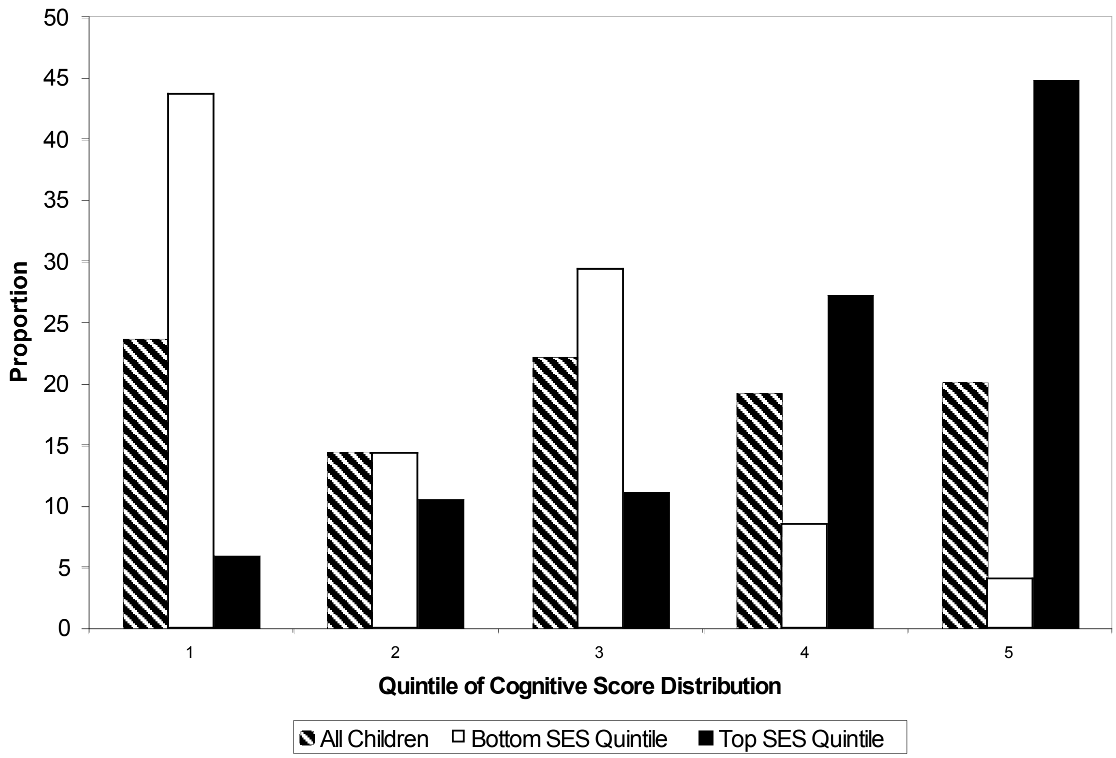
Source: Author’s calculations from the NLSCY.

Figure 11: Transition probabilities from the bottom quintile of the cognitive score distribution at ages 0-3 to the bottom quintile of the cognitive score distribution at ages 12-15, by quintile of family SES at ages 0-3



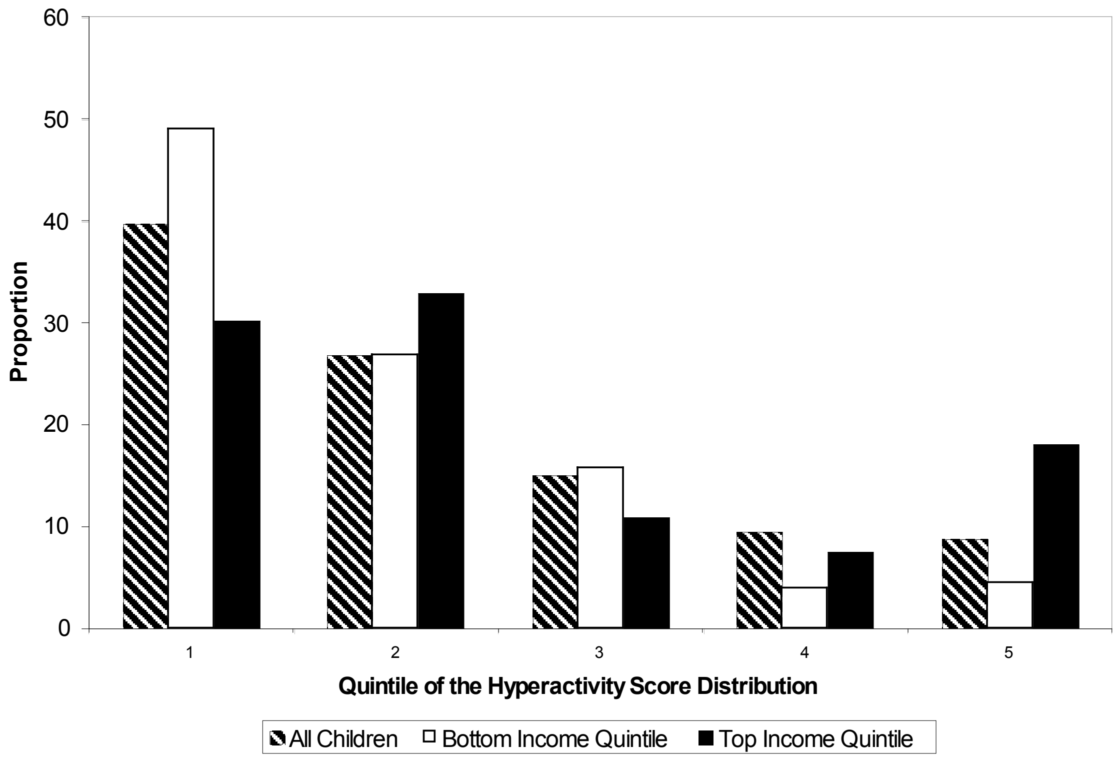
Source: Author's calculations from the NLSCY.

Figure 12: Transition probabilities from the top quintile of the cognitive score distribution at ages 0-3 to the quintiles of the cognitive score distribution at ages 12-15, by quintile of family SES at ages 0-3



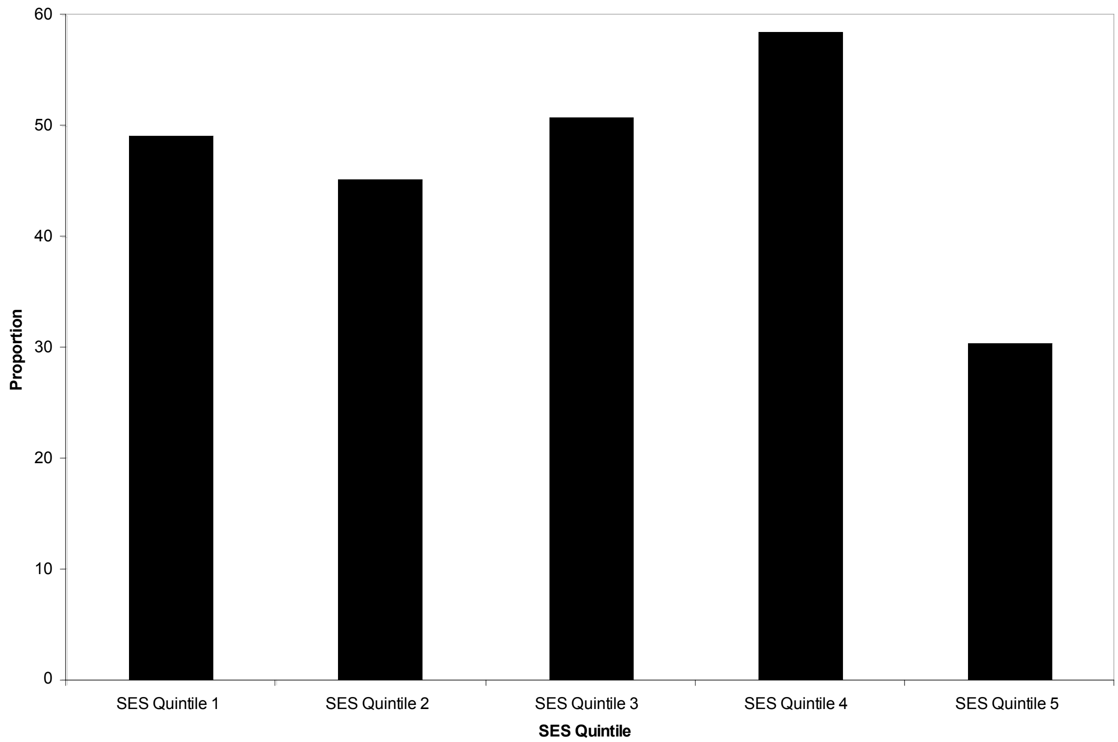
Source: Author's calculations from the NLSCY.

Figure 13: Transition probabilities from the bottom quintile of the hyperactivity scale distribution at ages 2-5 to the quintiles of the hyperactivity scale distribution at ages 7-11 by quintile of family SES at ages 2-5



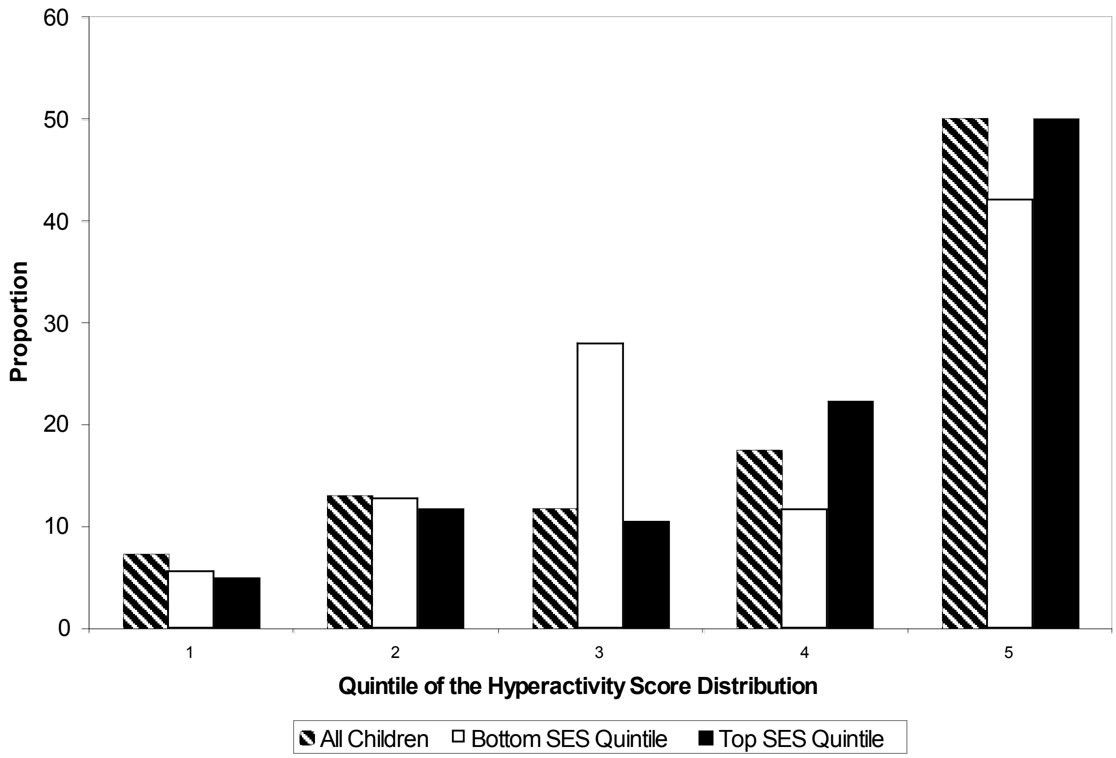
Source: Author's calculations from the NLSCY. Note the ordering of the aggression score is the reverse of the original data—a higher score means a better outcome.

Figure 14: Transition probabilities from the bottom quintile of the hyperactivity scale distribution at ages 2-5 to the bottom quintile of the hyperactivity scale distribution at ages 7-11, by quintile of family SES at ages 2-5



Source: Author's calculations from the NLSCY. Note the ordering of the aggression score is the reverse of the original data—a higher score means a better outcome.

Figure 15: Transition probabilities from the top quintile of the hyperactivity scale distribution at ages 2-5 to the quintiles of the hyperactivity Scale distribution at ages 7-11, by quintile of family SES at ages 2-5



Source: Author's calculations from the NLSCY. Note the ordering of the aggression score is the reverse of the original data—a higher score means a better outcome.

Table 1: Summary of studies investigating the impact of maternity/parental leave programs

Country	Outcome	Outcome Measured At Age	Increase in Care at Ages	Finding
Canada- Baker & Milligan (2010)	Motor/Social Development	Up to 24 months	6-9 months	No Effect
Canada- Baker & Milligan (2011)	Cognitive and Behavioural Development	4-5 years	6-9 months	Behaviour-No Effect Cognitive-Small, negative effect
Germany- Dustmann & Schonberg (2008)	Selective High School Wages	Teenagers	2-10 months	No Effect
Denmark- Rasmussan (2010)	High School Graduation GPA	Teenagers	4-5 months	No Effect
Sweden- Liu & Skans (2010)	High School grades and test scores	16 years	12-15 months	No Effect
Norway- Carneiro et al. (2010)	High School drop out	As old as 29 years	8-12 months	Reduction in High School dropout rate

Table 2: Summary of studies investigating the impact of public universal childcare programs

Country	Outcome	Outcome Measured At Age	Finding
Canada- Baker et al. (2008)	Behaviour	2-4 years	Negative effect
Canada- Lefebvre et al. (2008)	Cognitive-PPVT	4-5 years	Negative effect
Canada- Kottelenberg and Lehrer (2011)	Behaviour	2-4 years	Negative effect at mean Positive effect for low income children
Denmark- Gupta and Simonson (2010)	Non Cognitive	7 years	No effect for Preschool Negative effect of family daycare for low income males
Norway- Black et al. (2011)	Junior high academic performance	Grade 10	Positive effect: variation in daycare subsidy has no effect on use of care so impact may be due to additional income
Norway-Havnes and Mogstad (2009 & 2011)	Educational attainment, earnings	30-40 years	Positive effect: low income children main beneficiaries, negative for high income children

Table 3: Summary of studies investigating the impact of preschool programs

State	Outcome	Outcome Measured At Age	Finding
Oklahoma- Gormley and Gayer. (2005)	Cognitive, motor and language	4 years	Positive effect for Hispanics and blacks
Oklahoma- Gormley et al. (2005)	Cognitive	4 years	Positive effect for broader population
New Mexico- Hustedt et al. (2008)	Math, vocabulary and literacy	4 years	Positive effect – over sample of Native Americans and Hispanics relative to population
Georgia- Fitzpatrick (2009)	Cognitive	9 years	Positive effect for disadvantaged children in small towns and rural areas
MI, NJ, SC, OK, WV- Wong et al. (2007)	Vocabulary, math, print awareness	4 years	Positive effect for 8/14 outcomes- program in one state is universal
Lots of States- Cascio (2009)	High School dropout, Institutionalization	6-35 years	Positive effect for whites

Table 4: Estimates of the association of the cognitive and behavioural scores of children aged 0-5 in wave 1 and cognitive and behavioural scores at older ages

	Cognitive Scores				Behaviour Scores							
Score in Wave 1	MSD		PPVT		Hyperactivity		Anxiety		Aggression		Prosocial	
Wave 3												
Children's Age	3-7		7-9		4-10		4-10		4-10		4-10	
ScoreW1	0.163**	0.229**	0.152**	0.217**	0.412**	0.506**	0.402***	0.419***	0.261***	0.297***	0.260***	0.205***
	(0.024)	(0.061)	(0.039)	(0.070)	(0.027)	(0.050)	(0.019)	(0.036)	(0.012)	(0.022)	(0.012)	(0.023)
ScoreW1*Income		-0.001		-0.001		-0.002**		-0.0003		-0.0007**		0.0009***
		(0.001)		(0.001)		(0.001)		(0.0006)		(0.0003)		(0.0003)
Wave 5												
Children's Age	7-12		11-14		8-11		8-11		8-11		8-11	
ScoreW1	0.067**	0.067**	0.106**	0.020	0.271**	0.349**	0.201***	0.104*	0.136***	0.055*	0.144***	0.133
	(0.015)	(0.032)	(0.037)	(0.074)	(0.029)	(0.061)	(0.035)	(0.064)	(0.017)	(0.033)	(0.017)	(0.030)
ScoreW1*Income		-2.97e-06		0.002		-0.0014		0.0020*		0.0015***		0.0002
		(0.0005)		(0.001)		(0.0010)		(0.0011)		(0.0005)		(0.0004)
Wave 7												
Children's Age	11-16		16-18									
ScoreW1	0.090**	0.026	0.187**	0.334**								
	(0.025)	(0.053)	(0.048)	(0.108)								
ScoreW1*Income		0.0011		-0.003								
		(0.0008)		(0.002)								

Notes: The reported coefficients are from equation (1). Wave 1 MSD scores are for children aged 0-3, PPVT scores for children aged 4-5 and behavioural scores for children aged 2-5. W1~wave 1. Robust standard errors in parentheses. The cognitive scores are all standardized.

Table 5: Estimates of the association between behavioural scores at ages 2-5 and grade retention and high school dropout

	Hyperactivity		Anxiety		Aggression		Prosocial	
Wave 3 Outcome – Repeat a grade								
ScoreW1	0.001	0.005	-0.0002	0.003	-0.0003	-0.0005	0.0001	-0.0010
	(0.001)	(0.003)	(0.0013)	(0.003)	(0.0010)	(0.0021)	(0.0007)	(0.0015)
ScoreW1*Income		-0.0001		-0.0001		0.0000		0.0000
		(0.000)		(0.0000)		(0.0000)		(0.0000)
Wave 5 Outcome – Repeat a grade								
ScoreW1	0.002	0.0034	0.001	0.0002	-0.0013	-0.003	0.0001	-0.0020
	(0.002)	(0.0032)	(0.002)	(0.0031)	(0.0015)	(0.003)	(0.0008)	(0.0017)
ScoreW1*Income		-0.000		0.0000		0.0000		0.0000
		(0.000)		(0.000)		(0.0000)		(0.0000)
Wave 8 Outcome – High School Dropout								
ScoreW1	-0.001	0.0036	0.003	0.003	0.0004	-0.0006	0.0021	0.0022
	(0.003)	(0.0057)	(0.003)	(0.003)	(0.0026)	(0.0050)	(0.0016)	(0.0031)
ScoreW1*Income		-0.00008		0.012		0.0000		0.0000
		(0.00006)		(0.015)		(0.0001)		(0.0000)

Notes: The reported coefficients are from equation (1). W1~wave 1. Robust standard errors in parentheses.

Table 6: Estimates of the association between behavioural scores at ages 2-5 and math and problem solving scores at older ages

	Hyperactivity		Anxiety		Aggression		Prosocial	
Wave 3 Outcome – Standardized Math Score								
ScoreW1	-0.012	0.0007	0.002	-0.004	-0.010	-0.005	-0.005	-0.009
	(0.009)	(0.0141)	(0.013)	(0.023)	(0.010)	(0.017)	(0.007)	(0.012)
ScoreW1*Income		-0.0002		0.0001		-0.0001		0.0001
		(0.0002)		(0.0004)		(0.0002)		(0.0002)
Wave 5 Outcome – Standardized Math Score								
ScoreW1	-0.014**	-0.023*	-0.001	0.003	-0.004	0.010	-0.010**	-0.008
	(0.006)	(0.013)	(0.010)	(0.023)	(0.007)	(0.013)	(0.004)	(0.009)
ScoreW1*Income		0.0002		-0.0001		-0.0003		-0.0003
		(0.0002)		(0.0003)		(0.0002)		(0.0014
Wave 7 Outcome – Standardized Math Score								
ScoreW1	-0.016	-0.064**	-0.019	-0.081**	-0.025*	-0.029	-0.001	-0.008
	(0.011)	(0.023)	(0.019)	(0.036)	(0.013)	(0.024)	(0.008)	(0.015)
ScoreW1*Income		0.0008**		0.0012*		0.0001		0.0001
		(0.0004)		(0.0007)		(0.0004)		(0.0002)
Wave 7 Outcome – Problem Solving Score								
ScoreW1	-0.007	-0.049*	-0.006	-0.042	-0.002	-0.034	-0.019*	-0.039**
	(0.012)	(0.027)	(0.018)	(0.036)	(0.015)	(0.033)	(0.010)	(0.016)
ScoreW1*Income		0.0007*		0.0006		0.0005		0.0003**
		(0.0004)		(0.0006)		(0.0005)		(0.0002)

Notes: The reported coefficients are from equation (1). W1~wave 1. Robust standard errors in parentheses.

Table 7: OLS and sibling fixed effects estimates of the relationship between hyperactivity at ages 2-5 and math scores at older ages

	Full Sample	Siblings Sample		Full Sample	Siblings Sample	
Wave 3 Outcome – Standardized Math Score	OLS	OLS	FE	OLS	OLS	FE
Score1	-0.012	-0.018	-0.026	0.0007	-0.017	-0.046
	(0.009)	(0.020)	(0.021)	(0.0141)	(0.054)	(0.057)
Score1*Income				-0.0002	-0.0000	0.0004
				(0.0002)	(0.0000)	(0.0009)
Wave 5 Outcome – Standardized Math Score						
Score1	-0.014**	-0.015	-0.036**	-0.023*	-0.013	-0.029
	(0.006)	(0.011)	(0.015)	(0.013)	(0.020)	(0.032)
Score1*Income				0.0002	-0.0000	-0.0001
				(0.0002)	(0.0003)	(0.0006)

Notes: The reported coefficients are from equations (1) and (2). W1~wave 1, OLS~ordinary least squares and FE~fixed effects. Robust standard errors in parentheses. Standard errors for OLS estimates from "Siblings Sample" are clustered by mother.

Table 8: OLS and sibling fixed effects estimates of the relationship between Motor/Social Development at ages 0-3 and grade retention and high school dropout

	Full Sample	Siblings Sample		Full Sample	Siblings Sample	
Wave 3 Outcome – Repeat a grade	OLS	OLS	FE	OLS	OLS	FE
Score1	-0.012**	-0.061	-0.063	-0.026*	-0.137	-0.140
	(0.006)	(0.051)	(0.048)	(0.014)	(0.010)	(0.091)
Score1*Income				0.0003	0.0014*	0.0014
				(0.0002)	(0.0011)	(0.0009)
Wave 5 Outcome – Repeat a Grade						
Score1	-0.018**	-0.023*	-0.024	-0.027**	-0.016	-0.0023
	(0.006)	(0.014)	(0.018)	(0.010)	(0.021)	(0.024)
Score1*Income				0.0002	-0.0001	-0.0004
				(0.0002)	(0.0004)	(0.0211)
Wave 8 Outcome – Repeat a Grade						
Score1	-0.020**			-0.043**		
	(0.010)			(0.018)		
Score1*Income				0.0004**		
				(0.0002)		

Notes: The reported coefficients are from equations (1) and (2). Wave 1 MSD scores are standardized. W1~wave 1, OLS~ordinary least squares and FE~fixed effects. Robust standard errors in parentheses. Standard errors for OLS estimates from "Siblings Sample" are clustered by mother.

Table 9: OLS and sibling fixed effects estimates of the relationship between Peabody Picture Vocabulary Test Scores at ages 4-5 and grade retention and high school dropout

	Full Sample	Siblings Sample		Full Sample	Siblings Sample	
Wave 3 Outcome – Repeat a grade	OLS	OLS	FE	OLS	OLS	FE
Score1	-0.016**	-0.009**	-0.005	-0.034**	-0.020**	-0.008
	(0.003)	(0.004)	(0.004)	(0.007)	(0.009)	(0.006)
Score1*Income				0.0003**	0.0002**	0.0000
				(0.0001)	(0.0001)	(0.0001)
Wave 5 Outcome – Repeat a Grade						
Score1	-0.031**	-0.029*	-0.018**	-0.056**	-0.047**	-0.023
	(0.005)	(0.009)	(0.008)	(0.011)	(0.015)	(0.015)
Score1*Income				0.0004**	0.0003**	0.0001
				(0.0001)	(0.0002)	(0.0002)
Wave 8 Outcome – High School Dropout						
Score1	-0.022**			-0.045**		
	(0.009)			(0.018)		
Score1*Income				0.0004*		
				(0.0002)		

Notes: The reported coefficients are from equations (1) and (2). PPVT scores are standardized and are from waves 1-3 as described in the text. W1~wave 1, OLS~ordinary least squares and FE~fixed effects. Robust standard errors in parentheses. Standard errors for OLS estimates from "Siblings Sample" are clustered by mother.